2337

J. FRANCIS RUMMEL

An Introduction to Research Procedures in Education

SECOND EDITION

An Introduction to

Research Procedures in Education

EDUCATION FOR LIVING SERIES

Under the Editorship of H. H. Remmers

An Introduction to Research Procedures in Education

SECOND EDITION

J. FRANCIS RUMMEL

Professor of Education, University of Oregon

HARPER & ROW, PUBLISHERS

New York, Evanston, and London

LIBRARY OF CONGRESS CATALOG CARD NUMBER: *64–12797*

D-O

CONTENTS

TABLES

ILLUSTRATIONS

EDITOR'S INTRODUCTION

The first edition of a book is already in some respects obsolete when it appears in print. Particularly is this true when the content deals with research methodology in the behavioral sciences that undergird education and in education itself. In a world in which science and technology are making revolutionary changes, education and relevant research processes do not remain unaffected. After half a dozen years of a useful and successful book—to judge from its many adoptions—a revision is therefore clearly indicated. As never before, rational, scientific solution of educational problems is required with methods not beset by cultural lag.

Not only the novice in scientific research in education, but especially the consumers and school personnel, need to be aware of the ways in which empirical educational truth is arrived at. That they do not necessarily have such a meaningful, functional grasp of such matters I have again found in teaching in one of our great eastern universities where practicing public school teachers—earnest, sincere, but with little or no grasp of research methodology—were generally quite unable to read the technical literature. Professor Rummel's revision is well designed to fill this gap, written as it is to "bridge the gap between oversimplified accounts of scientific endeavors and advanced technical studies."

Within the frame of reference of the scientific method and the author's purpose, the plan and objectives of this revision could hardly change materially and has not done so. Its content has, however, been increased, sharpened in focus, and desirably expanded in certain areas, such as sampling, significance testing,

sociometric research, and the role of theory. Bibliographical references have also been updated. Perhaps not the least useful are the numerous examples from actual attempts at research of how *not* to do it.

It is a pleasure to present this revision of another significant book in the Education for Living Series that I have had the honor of editing for nearly a quarter of a century. Its mission was never more urgent.

H. H. REMMERS

PREFACE

This book is intended for use in the graduate departments of institutions of higher education, and is specifically directed to schools and colleges of education. It is designed as a guide to the scientific method which underlies all good research, and is written for all who are beginning their careers in education or in any other branch of the social sciences. It is not only aimed at the individual who is to be a *producer* of research, but also to those in larger numbers who are to be *consumers* of research. Both those who conduct research projects and those who read reported researches should know how and by what methods evidence is gathered, analyzed, and interpreted in the solution of problems. It should also be of value to committees as a guide for reflective thinking in the conception and solution of research projects. And it may serve field workers, graduate students, faculty members, and others who wish to study their problems objectively.

In presenting this text it is recognized that there is a substantial number of books, monographs, and other treatises on methods, techniques, and practices with which a student should become closely acquainted, and that no single publication will serve adequately as the only reference for a course in research procedures. However, the contents of this book provide the basic considerations and techniques in research methodology, with carefully selected references for more intensive study of each process and procedure included at the end of each chapter.

As a product of the field and the classroom, this book presents the nature of scientific research so that it may be clearly understood—the fundamental principles of problem solving are used

as the basic approach. It was written to meet the most general needs of the student rather than those of the instructor, who is presumed to understand research methodology. It seeks to bridge the gap between oversimplified accounts of scientific endeavors and advanced technical studies. With a minimum of technical terminology, and many simple illustrations, it introduces the student to the logical, psychological, and mathematical foundations upon which scientific investigations rest. Illustrative materials are, for the most part, drawn from the field of education, with excerpts from actual research proposals and projects.

Although the primary objectives and basic plan of this text have remained the same as in the first edition, the author has now combed the text thoroughly to make meanings more precise, to bring certain material up to date, and to expand some of the topics discussed. Notable is the clarification of the role of theory in conducting educational research. The chapter on development of the questionnaire has been expanded considerably by the inclusion of many illustrations of both good and poor techniques taken from actual research studies. The section on sociometric analyses has been enlarged to include more detailed explanations of sociometric processes and some of the newer techniques of analysis.

The text is organized into eleven chapters, covering fundamental considerations and techniques in research methodology, arranged in the order in which the stages of research normally arise: the development of the proposal to be submitted to the student's advisory committee, the processes of data collection and analysis, and finally, the writing of the report of findings and conclusions. The beginning of each chapter includes an outline of the major points of emphasis to facilitate study and review.

The first three chapters sketch the nature of research and introduce the student to the principles for the selection of a research problem and the planning and organization necessary to carry it out prior to the actual collection of data. Chapters IV through VII present techniques and illustrations for data col-

lection: the use of observation, interview, questionnaire, correspondence, and documentary analysis; while Chapter VIII introduces the neophyte to some of the basic designs for experimental types of studies, and Chapter IX presents various types of scaling procedures and sociometric analyses.

With the extensive development of data-processing equipment and techniques for the sorting, classification, and statistical computation of data in the past few years, an inclusion of procedures using punched cards is presented in Chapter X as an important feature of this book. This area of research procedures requires far more attention than it has received previously.

Chapter XI provides the beginning researcher with many guides for the writing of research reports. This is followed by Appendix A, "A Form Manual for Academic Research Reports," presenting some basic examples of the make-up of research reports, such as the proper form for footnotes, headings, tables, figures, and bibliographical entries. There are, of course, several acceptable ways to organize and present data; however, this manual of form presents only one, which is generally acceptable for the final preparation of typewritten reports. The student should find that this form manual provides answers to nearly all of the important questions which may arise in the writing of a research report.

Appendix B provides a review of the basic statistical concepts and computational procedures necessary in most research undertakings. It has been added to be used either as a basic topic to be included in a course in research methodology, or as a supplemental aid and review. Although it is not intended as a substitute for a comprehensive course in statistical procedures, it can provide an introduction to the basic concepts of statistics for those who have had little or no training in this area, and as a refresher for those who have not been working with statistical procedures recently.

While many volumes would be needed to cover all facets of research methodology, the author's intent has been to provide a brief introduction to those aspects of research pertinent to

beginning researchers which could be covered adequately in a one-semester course of study. He believes that he has come close to fulfilling his intent, although he has faced a serious problem of selection and compromise in presenting a text that will provide the basic knowledge and yet appear neither too costly nor too forbidding. However, brevity makes it possible for instructors to develop their own points of emphasis and to give greater attention to the specific problems of their own students.

The author wishes to express his gratitude to the various publishers who have granted him permission to use copyrighted materials. He also acknowledges with appreciation the many graduate students who have contributed to the development of this text through the identification of those problem areas of greatest concern to beginners in research.

J. F. RUMMEL

An Introduction to

Research Procedures in Education

CHAPTER I

The Nature of Research

Meanings of Research
- Definition
- Kinds of research investigations
- Individual and group research

Stages in the Development of Research
- Trial and error
- Authority and tradition
- Speculation and argumentation
- Hypothesis and experimentation

Universality of Research Methods
- Attempts to classify methods
- Good research uses a number of methods
- Rigid formulation of method impossible

Scientific Method of Research
- Steps in general research methodology
- Theoretical framework for research

Man appears to be universally beset by problems which must be met and solved. These problems are as varied as life itself. Scarcely a day goes by but that each of us does not investigate, study, and question some aspect of our environment to acquire facts which may help to solve problems that face us. Since most personal problems have been solved before, we might find answers to such problems if we read the right sources or asked

the right people. But the serious problems of society are difficult to solve as there are no sources or authorities who can provide us with a solution; the growing complexity of our civilization means that in the social sciences, and especially in education, we find that difficult new problems develop more rapidly than the old ones are solved. Since a problem is a doubtful case or question difficult of solution or settlement, it is necessary to understand the relevant facts in order to solve the problem. No reduction in vagueness can be made until the problem has been carefully defined and broken down into specific questions or subproblems. In all aspects of life—social, economic, educational, political, and business—there is increasing emphasis upon research to give man the factual data he needs to solve individual and social problems.

Research has other uses than solving problems. Man wants factual information to help him satisfy his curiosity and to learn the direction he is going, how far he has progressed, and how his progress compares with that of his fellows. Important as these types of research may be, this book will concentrate on research as a method of solving problems.

MEANINGS OF RESEARCH

Man uses various tools for the solution of his problems, some of which have been handed down to him from the past. In the use of some of these tools he may not question how the basic information was acquired or whether it has been verified. For example, man may passively accept custom or tradition, the dictates of various authorities, his own limited experiences, or syllogistic reasoning from propositions appearing to be self-evident. Finally, he may base the solution of his problem upon the findings of scientific inquiry, a synonym for research.

Definition

Research is a careful inquiry or examination to discover new information or relationships and to expand and to verify existing knowledge. It is the "manipulation of things, concepts, or symbols for the purpose of generalizing and to extend, correct,

or verify knowledge, whether that knowledge aids in the construction of a theory or in the practice of an art. The mechanic or physician is a research worker only when he attempts to generalize about all automobiles or all patients in a given class."[1] Certain writers on the topic have emphasized one or more characteristics of research that others have minimized, but the general nature of the activity is not in dispute.

Research discussed in this book will be "social science" rather than "natural science" oriented. There are many differences in experimental conditions between the two fields, principally because of the extreme difficulty in running a "control" experiment for purposes of comparison when the research topic is in a social science field.

Progress in research has been accelerated by the discovery and development of methodological procedures making the process more rigorous, discriminating, and dependable. The new electronic data-processing machines collate data and detect relationships so rapidly that it is possible to carry on investigations that would have been too costly and too time consuming only a few years ago. Methodological developments have been enhanced by researchers' dedicated belief that unbiased study of all relevant facts is the best way to solve many problems.

It must be admitted that the word "research" is abused in common speech in that it frequently is used to mean "looking up" something in a standard reference book and not the acquisition of new knowledge.

Kinds of Research Investigations

Research, as the typical college student understands the term, is a relatively long investigation of a properly limited topic carried out primarily in libraries with the results presented in a highly documented paper of some length. Of course the term *research* has broader and stricter meanings than this. In its wider sense it includes all specialized and thoroughgoing investigations in which educated people engage. In its specific use

[1] *Encyclopedia of the Social Sciences,* Vol. 13. New York: Macmillan, 1934, pp. 330–334.

it includes several kinds of advanced study and investigation which may be classified under four general categories, namely, library, life and physical science, social, and technological research.

Library research may be briefly described as a kind that is conducted primarily through the use of written materials most commonly located in large libraries. It is concerned with the seeking out of significant facts and interpretations from the past and from the extensive data and statistical information about contemporary life which are frequently found in government documents, professional journals, and similar sources. Studies concerned with the evolution of theories and research into possible cause and effect relationships are likely to rely heavily on the use of library material.

Library research that is worthy of the name necessitates generalizations and conclusions not previously appreciated. Unfortunately, it must be admitted that "for many a graduate student it is a euphemism for wholesale plagiarism."[2]

Life and physical science research is, for the most part, empirical. It tends to utilize laboratories more than libraries, and the resulting reports are often shorter than those based on written sources. Research of this type may be exemplified by the kinds of laboratory studies carried out in the fields of medicine, astronomy, geology, atomic energy, etc. It seeks, for example, new facts about the physical life of molecules, minerals, and men.

Social research is defined here to include research in both the social sciences and the humanities. It is devoted to a study of mankind in his social environment and is concerned with improving his understanding of social orders, groups, institutions, and ethics. This definition should be construed broadly enough to include research in such fields as foreign languages, philosophy, religion, etc. Social research is becoming increasingly important in the lives of people today.

Technological or applied research consists largely of the ap-

[2] Bergan Evans and Cornelia Evans, *A Dictionary of Contemporary American Usage*. New York: Random House, 1957, p. 420.

plication of the previously listed kinds of research to the immediate needs of industry, recreation, education, and the economic aspects of people today. This type of research has often been referred to as "applied" research rather than the "basic" type of research usually associated with the search for new knowledge in itself.

Individual and Group Research

This book is addressed primarily to the student who is first becoming acquainted with research as a major part of his academic work. Such students engage in research on an individual rather than a team basis. Much faculty research is also individual. In such cases, which include the great bulk of educational research conducted in universities and colleges, the choice of appropriate methods is made by the researcher on the basis of his own judgment, capacities, and interests. However, there is a growing tendency toward group or team research, research such as that conducted by a school system, educational association, or governmental agency. Educational research in colleges and universities which is financed by grants and contracts is increasingly on a group basis under a project director. In such cases several researchers assume responsibility for particular aspects of the project, and their activities are coordinated by the director. When the study design is planned by the group, a variety of methods may be suggested because each researcher views the problem and the way to solve it according to his own background, experiences, and training. This is one of the great advantages of the group approach.

STAGES IN THE DEVELOPMENT OF RESEARCH

The history of intellectual development has been characterized by forward spurts followed by plateaus of complacency. Throughout history man has evolved various approaches for answering perplexing problems about life. Even as early as Aristotle research findings and empirical knowledge were being used in the physical and biological sciences. Research in the social sciences developed much more slowly, probably because

the social sciences deal not only with topics which are less amenable to objective determination, but also because problems in the social sciences often involve strong vested interests that tend to make investigations proceed in an emotional atmosphere. Nevertheless, the research approach to problem solving has, in most disciplines, been preceded by three other approaches: (1) trial and error, (2) authority and tradition, and (3) speculation and argumentation.

Trial and Error

During the infancy of a science, observations are for the most part casual and qualitative—the sun rises, beats down strongly at midday, and sets; the moon grows from a crescent to full and then diminishes. In this first stage, man does not have logical explanations for all of the observed relationships composing a science, and he "muddles" along, trying one thing after another, until he finds an acceptable solution. As the process of sifting out those methods and procedures that do not produce satisfactory results continues, a few principles gradually emerge. Hence, sheer trial and error may be considered the first stage in the development of a science.

Authority and Tradition

In the second stage, "leaders" of the past are quoted. Often they were partly or completely wrong, but their opinions were stated with such assurance and force that they eventually became hallowed as traditional views. The development of the natural sciences has involved many clashes with tradition: the names of Galileo and Darwin are associated with especially bitter crossing of swords. Many propositions of religion and social action claim support from some sacred text, tradition, or tribunal whose decision on such questions is vested with finality. Political, economic, and educational questions are frequently determined by appeals to such authorities. People may rely on tradition if they lack the time or training to settle particular problems, and in some societies certain traditions and au-

thorities are considered so infallible that external force may be invoked to give sanction to their decisions.

Speculation and Argumentation

In the third stage the authorities are frequently doubted and solutions of fact are sought through debate. This is the stage of philosophizing, or speculation and argumentation. In the nineteenth century, Charles Darwin, the English scientist, extended his theory of natural selection and survival of the fittest to the conclusion that many varieties and species of animals could be traced back to common ancestors. A contemporary scientist in Switzerland, Louis Agassiz, had conducted extensive research on the description and classification of fish and opposed the theories of Darwin as being irreverent. Agassiz was one of a number of scientists who stood stoutly for separate acts of creation. These men stimulated many debates and discussions about the evolutionary theory of the origin of animals. Scientific men now, however, generally consider it quite as reverent to trace varieties and species to the working of great laws as to assert that the Creator established fixed types which defy the influence of time and season.

In the past, mankind has suffered because of attempts to solve scientific problems merely by means of reason and deductive logic. The result has led to such events as Aristotle's pronouncement on heavenly bodies and Galileo's recantation by command of "authorities." However, the solution of life's problems requires both a philosophical and a scientific attack. "Mankind will be best served while the philosopher remembers that facts are not established by dialectic and the scientist does not forget that the ultimate purposes of life are not a matter for objective determination." [3]

As soon as basic data are available in substantial quantity, speculation, instead of being based on *a priori* reasoning exclusively, becomes modified by empirical material. The more em-

[3] Veloruz Martz, "Philosophy and Science," *Encyclopedia of Educational Research*, New York: Macmillan, 1941, p. 797.

pirical material there is, the closer speculation becomes tied to reality.

Hypothesis and Experimentation

When students seek first to secure their facts and then draw conclusions, they have reached the stage of hypothesis and experimentation, a stage which may lead to a fifth and more precise stage if the information involved is capable of being reduced to quantitative terms. McCloy has defined this stage as a careful, extraordinarily earnest, and well-organized effort to ascertain essential facts about something at present unknown. He describes it as follows:

> It involves constructive, creative thinking, and scientific analysis based upon experimentation, historical documents, observation, collection of objective data, and careful ratings. It should be as mathematically precise as possible, as objective as the type of data permits, and subject to verification by others. The research worker should be absolutely impartial, unbiased, and intellectually honest. He should set out to ascertain the facts and seek the results with an absolutely open mind. He should not confuse opinion with fact. He must not reason from silence or from absence of evidence, but should consider these as historical background until verified.[4]

This fourth, the scientific or research stage of science, may conceivably be quite inadequate. As Martz has pointed out:

> . . . bare facts, objective data, never determine anything. They become significant only as interpreted in the light of accepted standards and assumptions, and these standards in the final analysis are not susceptible of scientific determination. . . . In ordinary life we seldom deal with bare facts but facts interpreted. This interpretation or evaluation is determined by the purpose to which we relate the facts.[5]

In considering this statement by Martz, the research worker should recognize that accepted standards may include not only objectively determined physical facts, but also may include atti-

[4] C. H. McCloy, *Techniques of Research in Physical Education.* Iowa City, Iowa: State University of Iowa, 1948, p. 1 (mimeographed bulletin).

[5] Martz, *op. cit.,* pp. 794–796.

tudes, values, and other less tangible factors. Even in school surveys, in which most of the data collected may be highly objective, the conclusions may be quite subjective. Good[6] has remarked that "such conclusions depend largely on committee deliberations or interpretations, with rather broad recommendations based largely on opinion or the impact of opinion on facts." While the problem of objectivity often perplexes survey experts and other investigators of status problems, too great devotion to a too narrow conception of scientific research should not center our attention on things that can be enumerated or counted to the general neglect of careful observation by skilled and experienced observers and of logical analysis as techniques for study of many important problems covered in survey investigations.

These various stages in the development of research are not clearly separated from each other; it has already been noted that scientific research may be interwoven with speculation and argumentation because the facts that have been ascertained must be interpreted. It is impossible to avoid value judgments in dealing with the interpretation of facts, because facts discovered are of no use unless they are applied to human problems. This necessitates criteria and assumptions, the choice and acceptance of which may bring us to a situation having some similarities to the authority and tradition stage.

UNIVERSALITY OF RESEARCH METHODS

The scope of research is broad, covering the whole spectrum of human interests. But no matter what the field, the researcher uses various "methods" of research; the classification of these "methods" has become so detailed and varied that a neophyte has just cause for confusion.

Attempts to Classify Methods

Several bases for classification are given in the following quotation:

[6] Carter V. Good, "Educational Research After Fifty Years," *The Phi Delta Kappan* (January, 1956), *37*:147.

Methods of research may be classified from many points of view: the *fields* to which applied: education, history, philosophy, psychology, biology, etc.; *purpose:* description, prediction, determination of causes, determination of status, etc.; *place* where it is conducted: in the field or in the laboratory; *application:* pure research or applied research; *data-gathering devices* employed: tests, rating scales, questionnaires, etc.; *character of the data* collected: objective, subjective, quantitative, qualitative, etc.; *symbols* employed in recording, describing, or treating results: mathematical symbols or language symbols; *forms of thinking:* deductive, inductive, etc.; *control of factors:* controlled and uncontrolled experimentation; *methods employed in establishing causal relationship:* agreement, difference, residues, and concomitant variation. Where there is a shifting of a point of view in a given classification, without warning or explanation, the result almost always is confusing; . . .[7]

Still another classification of research methodology, commonly encountered, is: historical, library, field survey, case study, statistical, genetical, and experimental.

A simple dichotomy used frequently in social science research is that between quantitative and nonquantitative (or descriptive and reflective). The increasing emphasis upon quantitative measurement of findings, the development of new data-processing equipment, and the improvements in statistical methods have resulted in a rapid increase in quantitative studies in behavioral sciences research during the past few years.

Good Research Uses a Number of Methods

Nearly all research projects require the use of more than one technique or method. Even though the principal portion of the project uses data obtained directly from the field, the initial phase of the report is usually a recapitulation of existing information about the problem and a description of its background. This may involve historical, library, and case study methods. The field work may result in the collection of several hundred questionnaires or interview schedules which are subjected to statistical analysis. A part of the study may lead the researcher

[7] Carter V. Good, A. Barr, and Douglas E. Scates, *The Methodology of Educational Research*. New York: Appleton-Century-Crofts, 1941, p. 207.

to conduct limited experiments to test certain hypotheses on conjectures that he has developed. These tests would be examples of the experimental method.

Rigid Formulation of Method Impossible

Since no two research undertakings, nor the researchers who conduct them, are exactly alike, it is impossible to set forth any rigid formulation of method or procedure. There is a wide variation in the conditions and circumstances which determine the objective nature of research projects in different fields. Thus, all methods defy portrayal in terms of formula or standardization. However, it is possible, taking into account the basic considerations and fundamental techniques of research, to outline in general how a research study should be conducted.

SCIENTIFIC METHOD OF RESEARCH

Scientific research may be defined simply as the systematic and refined use of specialized tools and procedures to obtain a more adequate solution to a problem than would be possible by less discriminating means. It starts with a problem, collects facts which are critically analyzed, and reaches decisions based on actual evidence. It may well involve tentative hypotheses and, on occasion, experimentation. It evolves from a genuine desire to know rather than from a wish to prove a point of view. As far as possible, it stresses a quantitative approach, seeking to know not only *what* but also *how much;* measurement is, therefore, an important aspect of scientific research.

Philosophers have frequently discussed the procedures involved in initiating an inquiry. All agree that initially the investigator should clear his mind of traditional viewpoints, but there are numerous differences of opinion as to what the next steps are. In writing about the steps in the thinking process, John Dewey[8] and Truman L. Kelley[9] have presented six steps which are analogous to the scientific method of research:

[8] John Dewey, *How We Think*. Boston: Heath, 1933, p. 12.
[9] Truman L. Kelley, *Scientific Method: Its Function in Research and in Education*. New York: Macmillan, 1932, p. 5.

1. *A felt need.* This may be considered as the occurrence of some felt difficulty in adaptation of means to a desired end, in identifying the character of an object, or in explaining an unexpected event.
2. *The problem.* Once one is aware of some question or problem or difficulty, the next step is to define it in terms of a problem statement.
3. *The hypothesis.* The third step is that of stating a possible solution for the problem. The solution may be based upon a hunch, a guess, an inference, or a theory.
4. *Collection of data as evidence.* The fourth step is the collection of data, information, or evidence to bear out the implications of a hypothesis.
5. *Concluding belief.* On the basis of the evidence the idea is corroborated or rejected and a concluding belief is formulated through the experimental analysis of the hypothesis.
6. *General value of the conclusion.* After a solution has been found to work, there is a mental "looking forward," the general purpose of which is to appraise this new solution in the light of future needs. This is the answer to the question, "So what?" that is often raised at the end of many research efforts.

Evaluatn

The first five steps, above, were set forth by Dewey and the final step was the contribution of Kelley as the crowning act in the process of reflection or thinking.

The point of view expressed in this text is that good research methodology must reflect good thinking, and that these steps in the thinking process might well serve as procedural guides in the development and execution of research investigations. The sixth step, however, as applied to research carried out by graduate students in institutions of higher education, has been somewhat controversial. Graduate students are usually required to develop research proposals and have them approved by individuals or faculty committees. Some or all of the members of a student's committee may believe that the researcher's responsibilities end before the sixth step. However, the view taken in this book is that the student should do more than merely unearth facts; he should evaluate them, come to a conclusion that involves taking a position on the solution of the problem with

which he has been dealing, and be willing and capable of defending this position in the larger context suggested by the sixth point. This view is discussed in greater detail in the following chapter which deals with the selection of a topic for research.

Steps in General Research Methodology

There are six steps in the development of a research project which have general applicability. While these steps are stated differently than those of the thinking process, it is obvious that there is considerable similarity in the two lists. Each of these steps is considered in detail in subsequent sections of this text and is merely listed at this time to provide a general outline for scientific research.

1. Selection of the topic or problem for investigation
2. Definition and differentiation of specific aspects of the topic
3. The framing of working hypotheses to facilitate the preparation of a logical study design
4. Collection of pertinent data
5. Analysis and interpretation of the data
6. Written report of the research study

Research Activities Conform to the Complete Process of Thinking

In research activities the researcher is faced at first with seeking answers to some nebulous problem. He then "spells out" this problem in definite form and breaks it down to specific questions for which answers are sought. To do these things, the investigator must provide himself with a working hypothesis or guess as to the probable results. He then goes to work to get enough facts to test his hypothesis. He sets up experimental procedures which he thinks will manipulate the persons or materials concerned to bring forth the necessary information. A schematic diagram showing the various steps in the research process, which parallel the steps of the "thinking process," is shown in Figure 1.[10]

[10] Adapted from Benjamin S. Bloom (Ed.), *Taxonomy of Educational Objectives*. New York: Longmans (David McKay & Co.), 1954, p. 107.

Experiments should be designed so that there are as few 2 uncontrolled variables as possible. There should also be control groups in which the variable factors are influenced in different ways or not at all. Many reference books in research methodology have indicated the desirability of reducing experiments to only one uncontrolled variable. This is desirable if it can be done. However, it is usually quite difficult, if not impossible, to carry out research in the social sciences and education without involving many factors or variables. Many of these are readily apparent, but many are either unknown or undetected. A researcher may think he has a single-variable experiment and make his analysis accordingly without knowing that he has been in error because he was not aware of the other variables that may have influenced his results.

In the popular novel, *Disputed Passage,* by Lloyd Douglas, the story unfolds in the classroom of Doctor Forrester on the opening day of State University. The course in anatomy, under the brilliant but irascible Forrester, was reputed to be the stiffest course in the entire four-year curriculum. As a part of his opening lecture he made the following remarks which have value for all students, and especially those who are interested in research:

> . . . while this anatomical laboratory, to which you are shortly to be introduced, is no place to look for fragrance or faith or fairy stories, it is at least honest, which can't be said of legislative halls or art galleries or cathedrals. You are here in quest of truth. Once a fact is amply attested you are to accept it, no matter how ugly it is; no matter how much you wish it wasn't so; no matter how violently it collides with what you have previously thought and would prefer to think. And don't make the mistake of imagining that the testimony is all in, and on file. Many a scientist, six feet underground, would suffer the agonies of the supposedly damned if he could come forth today and read some of his dogmatic remarks, long since reduced to utter nonsense by new findings. Remember that until a theory has been disproved—no matter how fantastic it may appear in the light of our current knowledge—it should be accorded the respect due to a proposition that might be proved—sometime.[11]

[11] Lloyd C. Douglas, *Disputed Passage.* New York: Grosset & Dunlap, 1939, p. 12.

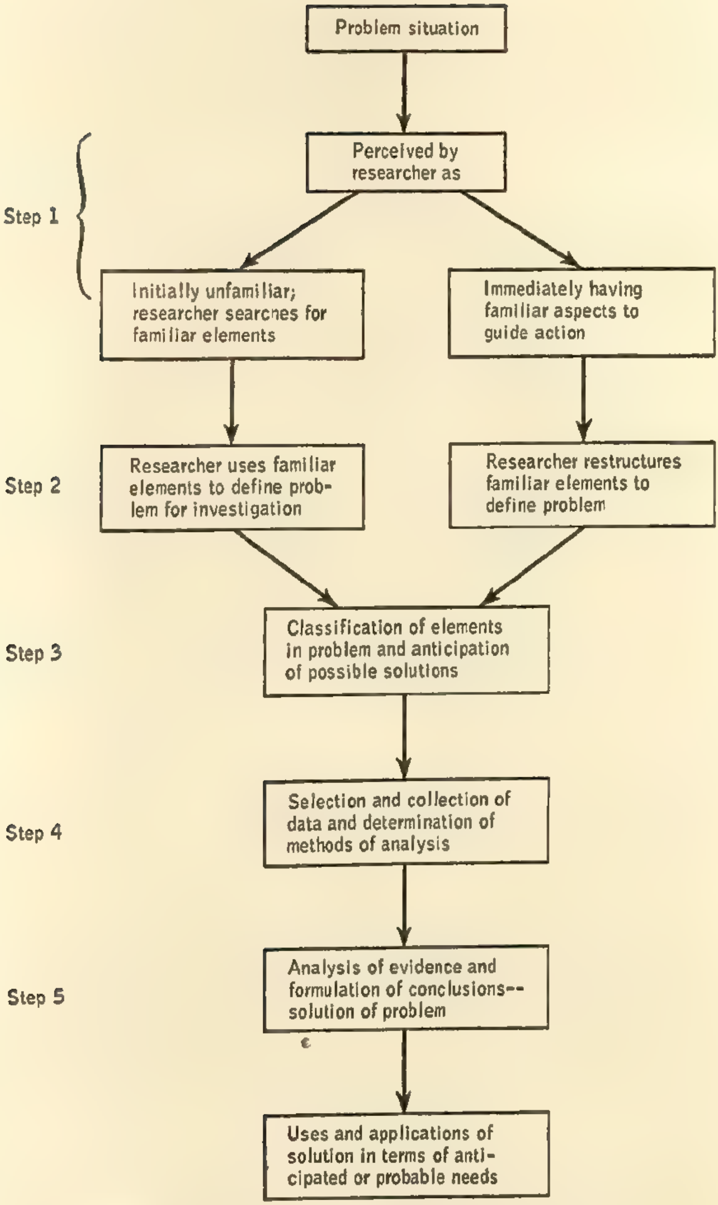


FIGURE 1. Schematic Diagram of the Research Process.

Theoretical Framework for Research

Basic to good scientific research is a *theory* which serves as a point of departure for the successful investigation of a problem. In this respect a theory is a tool of science since it may be used to define the kinds of data to be analyzed; it provides a guide to the way in which data are to be systematized, classified, or interrelated; it often points out new facts; and it often identifies areas in which our present knowledge is unsubstantiated or lacking entirely. In its simplest form a theory may be nothing more than a guess, a conjecture, a speculation, or an idea. A more complicated theory may be a summation of facts which have been accumulated in a given subject, an analysis of a set of facts in their ideal relationships to one another, a set of general or abstract principles, or a more or less plausible general principle offered to explain phenomena.

As more and more facts relevant to a theory are gathered, tentative generalizations can be made from them. These generalizations are usually referred to as a set of *postulates*. Deducing from a set of postulates one formulates a hypothesis—a statement capable of being tested and thereby verified or rejected.

> A hypothesis looks forward. It is a proposition which can be put to a test to determine its validity. It may seem contrary to, or in accord with, common sense. It may prove to be correct or incorrect. In any event, however, *it leads to an empirical test*. Whatever the outcome, the hypothesis is a question put in such a way that an answer of some kind can be forthcoming. It is an example of the organized skepticism of science, the refusal to accept any statement without empirical verification. Every worthwhile theory, then, permits the formulation of additional hypotheses. These, when tested, are either proved or disproved and in turn constitute further tests of the original theory. In either case they may be of use to existing theory and make possible the formulation of still other hypotheses.[12]

[12] William J. Goode and Paul K. Hatt, *Methods in Social Research*. New York: McGraw-Hill, 1952, pp. 56–57.

Because many theories are stated vaguely, do not specify the variables and conditions involved, and give inadequate cues as to what is to be measured, they do not form an adequate basis for research. Theories of a complicated and complex nature usually involve variables which we cannot measure or evaluate because instruments adequate for this are lacking at present. The more productive researches, as carried on by graduate students and reported in theses and dissertations, have referred to somewhat limited phenomena and have been stated in fairly simple terms. The role and application of theory as a framework for research is discussed in more detail in Chapter III.

SELECTED REFERENCES

The bibliography at the end of each chapter is included to provide students with a source of reference material to augment and enrich their understanding of research methodology.

Ackoff, Russell L., with the collaboration of Gupta, Shiv K., and Minas, J. Sayer. *Scientific Method: Optimizing Applied Research Decisions*. New York: Wiley, 1962. 464 pp. (This is an excellent mathematical approach to research.)

Dodd, Stuart C. "Scientific Methods in Human Relations," *The American Journal of Economics and Sociology*, 10:221-235, April, 1951. (Discussion of whether the scientific approach is the best or whether we should rely on tradition, appeal to the supernatural, or authorities. Several illustrations of the scientific method are presented.)

Enterline, H. G. "Summary of Needed Research in Business Education, Administration, and Supervision," *National Business Education Quarterly*, 20:48-52, May, 1952.

Gage, N. L. (Ed.) *Handbook of Research on Teaching*. Chicago: Rand McNally, 1963. (See chaps. 1, 2.)

Good, Carter V. *Introduction to Educational Research*. New York: Appleton-Century-Crofts, 1959. (Read especially chap. 1.)

Goode, William J., and Hatt, Paul K. *Methods in Social Research*. New York: McGraw-Hill, 1952. 386 pp. (See especially chaps. 1-6.)

Kerlinger, Fred N. "The Mythology of Educational Research: The Method Approach," *School and Society*, 88:149-151, March 26, 1960.

Mason, Robert E. "Philosophy of Education and Educational Research," *Journal of Educational Research, 50:*627–630, April, 1957.

Massialas, Byron G., and Smith, Frederick R. "Quality Research—A Goal for Every Teacher," *Phi Delta Kappan, 43:*253–256, March, 1962.

McCormick, Thomas C., and Francis, Roy G. *Methods of Research in the Behavioral Sciences.* New York: Harper & Row, 1958. 244 pp. (Read pp. 1–13.)

Mouly, George J. *The Science of Educational Research.* New York: American Book, 1963. (See chap. 3.)

Plochmann, George K. "The Three Layers of Research," *The Journal of General Education, 12:*170–175, July, 1959.

Punke, Harold H. "Mine is Research, Yours is Doubtful," *Peabody Journal of Education, 35:*13–22, July, 1957.

Selltiz, Claire; Jahoda, Marie; Deutsch, Morton; and Cook, Stuart W. *Research Methods in Social Relations.* Rev. ed. New York: Holt, Rinehart and Winston, 1959. 622 pp.

Shannon, J. R. "Traits of Research Workers," *Journal of Educational Research* (March, 1947), *40*:513–521. (Analyzes many of the traits of research workers. How do yours compare?)

Singleton, Carlton M. "Freedom to Research," *Elementary English, 38:*114–117, February, 1961.

Travers, Robert M. W. *An Introduction to Educational Research.* New York: Macmillan, 1958. 466 pp. (See especially chap. 1.)

Van Dalen, Deobold B., and Meyer, William J. *Understanding Educational Research.* New York: McGraw-Hill, 1962. (See chaps. 1–3.)

CHAPTER II

Choosing a Problem for Research

Academic Purposes of Specific Research Studies

- Term paper
- Field study
- Thesis
- Dissertation

How to Select a Problem Area

- The nature of a problem
- How to find a problem area
- Become a scholar in an area of specialization
- Read, listen, discuss, and think critically
- Follow up ideas that stem from present research
- Explore areas of dissatisfaction

Guiding Principles for Choosing a Topic

- High degree of personal interest
- Topic of significance
- Within the researcher's capabilities
- Necessary data are available to the researcher

Develop a Survey Bibliography for a Tentative Topic

- Sources of references
- Encyclopedia of educational research
- Review of educational research
- Graduate theses
- Books and Periodicals
- Biographical directories

The previous chapter dealt with the nature of research in general; this one first describes various ways to select a problem area for research and then presents some of the guiding principles for choosing a research topic. This book tends to emphasize empirical research as opposed to the mere review and analysis of previously written treatises. The primary reason for this is that empirical research more often calls into play a wider variety of techniques than does library research. There is no intention here to minimize the importance of research involving *a priori* reasoning or other nonempirical approaches; it is just that more purposes are better served by utilizing situations which necessitate the gathering and interpreting of original information.

ACADEMIC PURPOSES OF SPECIFIC RESEARCH STUDIES

In choosing problems for research, students should determine the purposes that are to be served. The scope of a problem, the methods of attacking it, and the manner of reporting vary somewhat according to how the results are to be used. The discussion in this chapter is directed to research conducted primarily by graduate students in institutions of higher education, and is more limited than it would be for research in general. Graduate research is usually reported in the form of a *term paper, field study, thesis,* or *dissertation.* Since the distinctions among these forms is oftentimes quite obscure, not only in the minds of students, but in the minds of faculty members as well, the following discussion may aid a student in selecting a topic for his purpose.

Term Paper

The term paper is usually considered the lowest level of these four forms of research reports in that it is not expected to be as extensive, nor as specific, as other kinds of research. Sometimes the topic is assigned by the instructor, sometimes the student is allowed to choose from a recommended list of topics the one in which he is most interested, and sometimes

he is allowed a free choice in the selection of his topic. Common purposes of the term paper are to provide a student an opportunity to explore a topic at greater length than it could be covered in the usual course of instruction; to reveal to an instructor a student's proficiency, or lack of it, in collecting, organizing, and reporting information in a logical manner; and/or to provide a partial base upon which a student may be given a grade for a course.

Frequently, in a term paper a student reports everything he could learn about his subject, and he is not required to seek for knowledge not already available. Due to a lack of time, or background, and of access to the more scholarly works, a truly comprehensive treatment and discussion of a topic is almost impossible. However, it should not be a rewording or rehashing of articles or reference books. "It should not be the stringing together of quotations from several authorities interspersed with undocumented paraphrases from other authors. It should be a document which is written by a student who has searched with intelligence through varied sources for certain facts which he recognizes as essential to his chosen subject."[1]

Field Study

The field study is a step above the term paper in its level of research, in that it is not written in partial fulfillment of the requirements of a *single* course of instruction, but in connection with an area of training or specialization for which a student is presumed to have had *several* courses. The topic of a field study may be one of highly practical value to a student in preparing for some specific vocational objective. It does not necessarily have to be original or to seek for new knowledge of general importance. It may be a study of a specific operation in a specific setting. However, it should have the characteristics of individual investigation; it should involve the collection of data, analysis of the data, documentation of citations, and sound conclusions; and it should be presented in a scholarly report.

[1] Lucyle Hook and Mary V. Gaver, *The Research Paper*. Englewood Cliffs, N.J.: Prentice-Hall, 1948, p. 1.

Thesis

A thesis differs somewhat from a field study in that it should establish a point that has hitherto been unsuspected or unsupported. It should attempt to discover procedures, rules, and principles relating to the various aspects of education. It should give a student training in the art and techniques of scientific inquiry and should have an important effect in determining the pattern of his subsequent interests and modes of thinking.[2]

According to Albaugh, another important difference between a thesis and a field study or term paper is the matter of audience.

> Frequently, a term paper [or field study] is seen only by the writer and by the person who grades it; occasionally it is seen by the writer's classmates; but in either event, once the term is ended, the paper ordinarily is forgotten. A thesis, on the other hand, may be seen by anyone who cares to look at it, and it can never be completely forgotten. Once the final draft is typed and submitted to the faculty, the thesis becomes the permanent property of the institution granting the degree. Bound and shelved in the library, it stands for an eternal monument to the writer's ability—or lack of ability. Furthermore, many colleges publish abstracts, and if a student or a faculty member at another institution decides, upon reading the abstract, that he would like to read the thesis, he may borrow the thesis through interlibrary loan, in which case the writer's reputation is not confined to the campus upon which he pursued his graduate work.[3]

Dissertation

The doctoral dissertation represents the highest level of research carried on by graduate students. It not only serves the function of being a very rigorous examination of a student's ability to conduct scholarly research, but, as Long has so ably stated:

[2] John A. Long, *Conducting and Reporting Research in Education*. University of Toronto, Department of Educational Research, Toronto, Canada: 1936, Bulletin No. 6, p. 2.

[3] Ralph M. Albaugh, *Thesis Writing*. Paterson, N.J.: Littlefield, Adams & Co., 1951, pp. 5–6.

It is a test of the student's scholarship, not so much in the direction of his ability to reproduce as in the direction of his capacity to create. The student must not expect an institution to place on him the seal of higher scholarship as a reward for any amount of unimaginative plodding. Ambition, industry, perseverance, and honesty of purpose are very commendable traits, but they are not enough; an advanced degree cannot be granted merely for prolonged effort. The dissertation must furnish evidence not only that the candidate is capable of sustained application in the solution of a problem, but also that he is a person of imagination, that he possesses initiative and originality to a marked degree, that he is a master of those techniques appropriate for his problem's solution, that he is gifted with a capacity for objectivity in investigation, and that he has the ability to report his study in an acceptable manner.[4]

The terms *thesis* and *dissertation* are oftentimes rather loosely used; but more precisely, a *thesis* is the research report required in partial fulfillment of the Master's degree and a *dissertation* is the final research report required for the Doctor's degree. While there is much similarity between them with respect to individuality of investigation, the collection and analysis of data, and the formulation of valid conclusions, they differ decidedly in scope of the topic for investigation. The dissertation topic should be much broader, the treatment of it should be more exhaustive, and the conclusions should permit wider generalizations than a topic acceptable for a thesis.

HOW TO SELECT A PROBLEM AREA

When starting a research project, one should first determine its suitability in terms of both external and personal factors as actual work on a research problem should not begin until its value and feasibility have been appraised. Of course, the problem must first be chosen before it can be assessed, and this is often the most difficult step for the graduate student. The remainder of this chapter will discuss the way to select a general problem area and to choose a specific topic within that area.

[4] Long, *op. cit.*, p. 3.

The Nature of a Problem

A thesis or dissertation should have a problem as its topic. It should not be solely the tabulation of answers to a questionnaire sent out, or a summation of answers to identical questions obtained by interviews. Neither should it be a mere recitation of facts, even though the facts may be historical and not previously available. A thesis or dissertation *must* include an analysis of the factual material collected and present one or more conclusions based on this analysis.

For example, a thesis or dissertation may report an investigation of the forces which shape certain community attitudes toward the public school system. To meet the standards set forth in this section, it would have to be more than a mere presentation of different attitudes and different forces at work in various communities. What are the interacting factors present in a community? To what extent do they affect individual or group attitudes? What are the relationships among various reference groups and their attitudes toward the school system? What are the relationships among social class, demographic characteristics, and attitudes toward public schools? If the data are such that there is opportunity for analysis and conclusions, the topic can be stated in the form of a problem.

It has been said that "in some sense, the basic characteristic of a problem is the presence of a hiatus between two thoughtways."[5] If, for example, actual developments are either inconsistent with existing theory, or not taken into account by any theory, then there is a "hiatus" between present developments and our understanding of them. Such a situation constitutes a "problem."

Research problems, and the generalizations resulting from their study, vary as to the degree of their abstraction. A study of the actual costs of operating school busses over various routes to determine the most efficient bus scheduling and routing will be more empirical than one evaluating the edu-

[5] Thomas C. McCormick and Roy G. Francis, *Methods of Research in the Behavioral Sciences*. New York: Harper & Row, 1958, p. 16.

cational programs of schools operated under different philosophies of education. However, the more work that is done in a particular field, the more abstract the problems can become. The first problems to be tackled will usually be of a very immediate nature, but as these are solved others will be exposed which will tend to be more abstract. In many fields, the accumulation of information will eventually make it possible to express certain relationships in terms of mathematical functions. Once this stage has been arrived at, both the problems presented and the conclusions drawn will be more abstract than descriptive.

How to Find a Problem Area

One of the more frustrating experiences of a graduate student is hunting for a research problem. He is aware of the large number of research studies turned out every year and wonders if there is anything left that needs investigating. But the very process of research exposes more and more problems for research. As the frontiers of knowledge are expanded, a greater area of unexplored phenomena is revealed. Therefore there is no lack of problems to be investigated, but a graduate student may think there is because he does not yet have the ability to recognize suitable topics. The following ways of finding or recognizing a research problem are only suggestive; good topics in the field of education are not easily found, but orderly methods can facilitate the search.

Become a Scholar in an Area of Specialization

The best way to find problems is to become a scholar in one or more specialties as soon as possible. This will enable one to analyze the field in which he is interested into its separate parts and to develop an understanding of the accomplishments of completed research and of the problems yet unsolved. Look for gaps or deficiencies in explanations. Watch for inconsistencies and contradictions. Observe the problems that grow out of a procedure in which you are interested and competent. Read the bibliographies of recent research in the field. A thorough

mastery of what is already known in an area is the most fruitful way of finding out what is lacking or needs to be explored.

One comprehensive lead to problems needing research in the field of education is the report of the joint committee of the American Association of Colleges for Teacher Education and the American Educational Research Association.[6] Between 1951 and 1954 the Committee determined upon a procedure to identify the categories or specific areas in the field of teacher education and to aim toward a compilation of significant and timely research titles within each of these categories. This process resulted in the identification by the Committee of five general areas with approximately five categories under each. A total of 568 suggested research topics were presented in the general areas as follows:

116—Scope, function, and objectives of teacher education.
161—Organization and administration of teacher education.
90—Student personnel programs and services in teacher education.
114—Curriculum and instruction, including laboratory experiences.
87—Relationships of teacher education institutions with other professional and public groups and agencies.

Each of these areas is preceded by a brief expository paragraph, and each category subdivision is preceded by a similar introduction to point up especially the specific areas for research needed on problems falling within the category.

Read, Listen, Discuss, and Think Critically

In reading, in discussions, at lectures, and in informal conversations, acquire the attitude of questioning every concept and procedure in education and seek supporting evidence. Follow clues and suggestions obtained in this manner. Develop a habit of carrying around a notebook in which to jot down ideas as they come to you. Read the best studies in your field. Criticize and challenge statements made in professional periodicals, in books, and in research studies.

[6] Roben J. Maaske (Chairman), *Needed Research in Teacher Education.* Oneonta, New York: American Association of Colleges for Teacher Education, 1954. AACTE Study Series, No. 2, 62 pp.

Associate with people in research and confer with faculty members in your area of interest. Seek the company of the intellectually competent individual who likes to argue or debate the opposite side of any question. Seek discussion with those who differ with your own basic beliefs in your field. They may be wrong, but they may provide useful leads in your search for a topic. At all times work and think critically.

Follow Up Ideas That Stem from Present Research

A number of suggestions of researches needed are to be found in the reports of completed research studies. In carrying out a research study, the researcher invariably finds that many minor problems, and even some problems of great importance, are discovered in pursuing the original line of investigation. Many times a doctoral dissertation includes a section on needed research that is revealed by the study.

In other instances, many studies have stopped with tentative conclusions due to an unforeseen lack of data or an inability of the researcher to analyze the data for conclusive results. In addition, some studies, conclusive for a limited sample of data, as, for example, a study of the efficiency of certain methods of instruction within a single classroom, may be repeated with more extensive data to reach important conclusions, generalizations, or applications.

In some cases it is desirable to repeat previous researches to verify the findings. In other cases, it may be necessary and desirable to duplicate studies under various conditions to be assured that all factors that were considered to be controlled in the original study were, in fact, controlled. For example, a study made in 1940 of teacher certification standards and teacher training programs might be done again in 1965 to see what changes had occurred in the interval. This would give the researcher an opportunity to analyze why changes had or had not occurred.

Sometimes completed studies offer only tentative conclusions because of an unforeseen lack of data or the researcher's inability to analyze the data conclusively. For example, a pioneer

study in a field may have utilized as much statistical data as the researcher had the time and capability to handle, but it may now be possible, using electronic data-processing equipment, to undertake a much more exhaustive study which may either verify or modify the conclusions given in the original one. In addition, some studies are only conclusive for limited geographical areas; conducting identical studies in other areas where income, education, and other socioeconomic factors are different may be of real value.

As this section has suggested, changes occur due to the passage of time which can affect the conclusions of earlier studies. Business studies are seldom conducted in laboratory situations, so that the introduction of new causes and the disappearance of old ones is likely to affect results. A comparison between present and past causes provides a method of estimating certain cause and effect relationships. The researcher can also attempt to isolate the effect of certain causes by conducting similar studies at approximately the same time in slightly different situations, or he may wish to learn if all the conditions that were assumed to have been controlled in a former study were, in fact, controlled.

Explore Areas of Dissatisfaction

If a student were to listen to the expressed likes and dislikes of the people around him, he would soon notice that many more dislikes were being expressed than likes. People have a tendency to express their dissatisfactions. Why are these people dissatisfied? What reasons do they give? With what are you dissatisfied in your field? What problems does it suggest? Investigate these.

GUIDING PRINCIPLES IN THE CHOICE OF A TOPIC

Once the student has selected a problem area, he may already have observed a number of possible specific topics which should be written down. The next step is to select that one upon which he will focus. This process may be quite time-consuming because all too often a student begins working on his selected thesis subject only to discover, once he is well into his project,

that it is better adapted to a term paper. On the other hand, students may choose topics so broad that the written report would run several printed volumes in length.

The topic the student selects from among the many possible ones should meet certain requirements. No major effort should be expended until the researcher is convinced that his topic does meet these. The literature of educational research is replete with criteria for the selection of a problem and standards for its evaluation. These are reviewed at length by Whitney,[7] Good and Scates,[8] and Perdew,[9] but they may be summarized into four principles, in terms of personal interest, personal capabilities, value of the topic, and availability of data.

High Degree of Personal Interest

The first, but not necessarily the most important, requirement of a research topic is that it be of interest to the student. Educational psychologists generally contend that little learning occurs without a high degree of interest. Similarly, little research is carried to completion without the spur of the researcher's internal drive. A student should ask himself the following questions:

1. Have I merely a passive curiosity in the solution of the problem I am considering, or is it a truly consuming intellectual drive?
2. Does my interest exist solely because of anticipated rewards such as obtaining an advanced degree or financial gain?
3. Is my interest in a particular solution of the problem so strong that it might bias my analysis?

The more facets of a problem which interest the student, the greater will be his desire to seek its solution. On the other hand, care must be exercised not to confuse interest in the problem with an overwhelming desire for one particular answer; a person should not engage in research only to "prove" a prior

[7] Frederick L. Whitney, *The Elements of Research*. Englewood Cliffs, N.J.: Prentice-Hall, 1950, pp. 80–94.

[8] Carter V. Good, and Douglas E. Scates, *Methods of Research*. New York: Appleton-Century-Crofts, 1954, pp. 33–102.

[9] Philip W. Perdew, "Criteria of Research in Educational History," *Journal of Educational Research* (November, 1950), *44*:217–223.

bias, for then he will tend to look only at those findings that support his view. If the student selects a topic of great interest to him, he must be willing to interpret his data and arrive at his conclusions without prejudice, regardless of whether they agree or disagree with any preconceived or contemplated outcome.

Topic of Significance

The significance of a topic deals with its novelty, timeliness, and academic and practical values. One might consider the following questions in determining the value of a topic:

1. Is it likely that the results of the study will add to the present body of knowledge, or will it only duplicate what has already been done?
2. Does the field need reworking?
3. Are there gaps in verified knowledge that need to be filled?
4. Are the results that may be obtained of practical value to business, society, government, or other agencies?
5. Who might be interested in the results?

To avoid unnecessary duplication, a student should check thoroughly the professional literature in the field of his topic to determine what has been done previously. In some cases, however, the field will need reworking due to changes in factors and conditions that were present at the time of a previous investigation. This is especially pertinent with respect to survey-type studies. On the other hand, a student should not be reluctant to conduct research on topics that have been investigated previously if there is any doubt as to the validity of the conclusions or the appropriateness of the methods and procedures of research used. Seldom are the results of research, especially in the social sciences, absolutely conclusive. "The attack of the same tentative generalization reported by another worker in another field, with new subjects and material, with refined techniques, and by better methods, is highly scientific."[10]

If a topic can make an application of knowledge to practical,

[10] Whitney, *op. cit.*, p. 90.

everyday procedures, it deserves scientific recognition. Spence[11] has said that all research is designed for eventual practice. So-called "pure" research is carried on with the belief that the resulting knowledge will lead at some time to a better world and the researcher of this type is not asked to justify his work in terms of immediate social problems. Spence also states that research workers can use information in connection with practical problems and at the same time increase the effectiveness of research. People often fail to realize the relationship between the technical developments of modern civilization and a long learning period. There are many situations in which people have technical knowledge but do not have the background necessary to use such knowledge. Coupled with this lack of know-how is the lack of a desire to change, which often comes from fear of insecurity.

The consumers of research findings should also be considered. Other things being equal, research revealing results that are desired by a large body of individuals has importance. The adherence to pure academic research, without thought of who might be interested in it, has frequently brought forth such satirical criticisms as the following:

> Every year about a thousand young men and women go off into justly-neglected corners of knowledge and assemble tiny scraps of more or less useless information into a little pile of dust, which, adorned with comparative tables, correlative graphs, and other forms of academic parsley, is served up as a thesis. The reward is the title of Doctor of Philosophy, which enables its recipient to ascend the educational ladder and in time teach other young men and women how to scrape together their own heaps of dust or doctoral dissertations, upon such themes as the Measurement of Typical Pessimistic Attitudes Among the Ducks of Lake Cayuga Region, or the Use of the Personal Pronoun in the Posthumous Works of Miss Atlantic City of 1936. (Author and source not known).[12]

Research does not necessarily have to have an immediate

[11] Ralph B. Spence, "Research and Practice: The Case for Impure Research," *Growing Points in Educational Research.* Washington, D.C.: American Educational Research Association, 1949, pp. 297–300.

[12] Hugh B. Wood, *Compilation of "Sense and Nonsense."* Eugene, Ore.: University of Oregon. *Curriculum Bulletin* (December 20, 1951). No. 100.

practical value. What may appear to be just a high-level academic exercise may at some time be recognized for its contribution to applied knowledge. In February, 1945, a United States Army Intelligence colonel walked into the Military Geology Branch of the U.S. Geological Survey with about a quart of sand which had been used as ballast from one of a number of Japanese balloons that had fallen in the Pacific Northwest, causing considerable incendiary damage. The military were anxious to know where the balloons had been coming from and appointed four scientists to analyze the sand. According to three micropaleontologists and one petrographer, the sand came from some sheltered beach. Its source rocks were both metamorphic and volcanic, but not granite, and it came from no beaches on the American coast or the nearer islands, for it contained organisms that did not appear in these sands and was lacking in other organisms that should appear.

Library research finally turned up a research report written by two French geologists in 1889 on the "Diatomées Fossiles du Japon." In this report were pictures of the same kind of fossils found in the sample of sand. By the use of Japanese geologic maps and papers prepared by Japanese scientists many years earlier, the Survey scientists finally narrowed the search for the source of the balloons to two beaches near Tokyo. When the war was over, the accuracy of their efforts was verified. "The balloon job gave the geologists a particular sense of satisfaction, for it proved to them what they have long contended—that no probing into the secrets of how the earth is made is ever wasted. Someday, somewhere, somehow, even the most abstruse and seemingly useless bit of scientific knowledge—even a fifty-six-year-old study of Japanese sand—will prove to be of practical value to somebody.[13]

Within the Researcher's Capabilities

The student should consider his own personal qualifications in background of training, physical and financial resources, and

[13] Harold H. Martin, "Uncle Sam's Treasure Hunters." *Saturday Evening Post* (July 17, 1954), *27*:22 ff.

available time in choosing his topic. He might ask himself the following questions:

1. Am I equipped, or can I equip myself, with the skills, abilities, and specialized knowledge required to attack this problem?
2. Is the topic feasible within the limits of my time and finances?
3. Is the topic one for which I can obtain a dissertation or thesis faculty advisor and committee as well as the proper technical help, such as library reference materials, that I will need?

Many a graduate student has embarked upon a research project only to find that when he has reached the point of making analyses of his data that he lacked the necessary statistical skills or educational background to continue his study satisfactorily. In selecting a topic, the student should attempt to define it to the point where he can determine the kinds of data necessary and the methods of treating the data. Then he should consider whether or not he has the requisite abilities to continue at that time or whether he can take the necessary time to develop them. If he is lacking in those skills and cannot, within the time available to complete his work, develop the necessary skills, he might just as well abandon the topic and seek another within his capabilities.

At times students have proposed topics for research which no one in the institution in which he wishes to carry on the research is willing to sponsor. The student should determine the availability of a sponsor for the entire time he plans for his research and the completion of his graduate program. In some instances, students have proposed topics for which no faculty member in the institution believed himself qualified to act as the sponsor. Good has pointed out many factors which should be considered in choosing a research adviser:

Leave of absence, a heavy teaching schedule, an already excessive number of advisees, concentration on writing or research, numerous speaking engagements, ill health, or personality difficulties on the part of a particular professor may render him relatively unavailable for additional assignments to the extent that the graduate student

may wish to turn elsewhere for thesis guidance or at least to consider the hazards involved in securing the necessary conferences and advice.[14]

Some topics are such that the institution in which they are proposed will not permit their investigation. Topics that deal with some moral issues, or involve comparisons among personnel within the institutions, or comparisons among various institutions, or those that might adversely affect public relations are of the type for which institution approval is often difficult and sometimes impossible to obtain. A student should discuss a proposed topic with an adviser or head of his department to ascertain whether or not administrative cooperation would be available prior to beginning detailed work on it.

Necessary Data Are Available to the Researcher

No matter how well a topic can meet the foregoing criteria, it cannot be solved unless the researcher can obtain the necessary data for its investigation. Some studies need to be based upon interviews with people impossible for the student to contact because of geographic inaccessibility, language barriers, social restrictions, or racial conflicts and prejudices. Some studies are stymied for lack of information because the necessary data are classified as "top secret." Some have to be rejected because of lack of satisfactory controls in the handling of the data or for lack of adequately developed techniques or instruments. A student should investigate the availability of data to obtain satisfactory answers to the following questions:

1. Are the data necessary to the solution of the problem readily accessible to me?
2. Are techniques necessary for the collection and analysis of the data developed to the point that the conclusions will not have to be so qualified as to make them of little or no value?
3. If certain classes of data are unavailable, can the problem be limited in scope or revised so that a satisfactory treatment can be planned using the available information?

[14] Carter V. Good, and Douglas E. Scates, *op. cit.*, p. 66.

DEVELOP A SURVEY BIBLIOGRAPHY FOR A TENTATIVE TOPIC

A survey bibliography should be compiled as an essential step in the selection and definition of a problem. It enables one to avoid unnecessary duplication of another's work and it provides many suggestions as to method, formulation of a problem, and ideas for other problems. One usually begins his search with a somewhat nebulous idea of a topic for research. By an organized program of reading the literature pertaining to the nebulous idea, he can begin to sharpen up his definition of the topic. This, in turn, provides the basis for a more specific search of the literature which further clarifies the problem. Thus, through a reciprocating sort of procedure of formulation-reading-and-reformulation, a problem develops to a point of definition.

Sources of References

To make a systematic canvass of the literature pertaining to a problem, the most fruitful lead is, first, to locate the most complete bibliography in the field of interest. For example, if one wanted to study a problem dealing with teacher competence, an investigation of the December, 1950, issue of the *Journal of Experimental Education* would provide an excellent point of departure. That issue was devoted to an annotated bibliography of over 1000 studies in the field of teacher competence. The difficulty confronting most students, however, is in locating similar bibliographies for their particular studies. The following sources of references are suggested as being the most helpful to students in education, although they represent only a few of the number of sources available.

Encyclopedia of Educational Research[15]

This book gives the outstanding research on a large number of educational topics, the specific problems considered, their

[15] Walter S. Monroe (Ed.), *Encyclopedia of Educational Research.* New York: Macmillan, 1941 and 1950, C. W. Harris (Ed.), 1960.

findings, and the various issues studied or needing further study. Most of the articles were written by outstanding specialists in their particular fields. Three editions, published in 1941, 1950 and 1960, are available. No student should start work on an educational topic until he has reviewed this publication.

Review of Educational Research

This particular journal of the American Educational Research Association has been published five times a year since 1931. As a rule, it covers each of fifteen major subdivisions of education in a three-year cycle. The first treatment of each topic reports the researches on it to date. Each subsequent treatment includes the pertinent research for its past three-year period. This publication does not describe extensively researches on a topic, but it does indicate the major problems and findings and provides numerous references and bibliographies. Each issue lists, on the inside of the back cover, the various areas covered in the series.

Graduate Theses

There have been a number of guides to Masters' and Doctors' theses published in education. Some of them have been rather sporadic in their publication schedules and are difficult to locate. A few of the readily available sources are as follows:

United States Library of Congress. *A List of American Doctoral Disertations Printed in 1922.* Washington, D.C.: U.S. Government Printing Office, 1933. (This is an example of the annual list published since 1912.)

Bibliography of Research Studies in Education. United States Office of Education Bulletin. No. –, (Year). Washington, D.C.: U.S. Government Printing Office. (Annual bulletins from 1926 to 1940. Suspended during World War II.)

Journal of Educational Research. (Doctoral dissertations under way were listed annually, usually in January, from 1931 to 1946.)

The Phi Delta Kappan. (Doctoral dissertations under way and completed were listed annually, usually in a spring issue, from 1949 to 1951.)

Research Studies in Education—A Subject Index. (Compiled by Mary Louise Lyda and Stanley R. Brown as a Phi Delta Kappa National Project, 1952 to 1953, with yearly supplements planned.) Boulder, Colo.: Phi Delta Kappa, University of Colorado.

Dissertation Abstracts. Ann Arbor, Mich.: University of Microfilms, Monthy, 1938–date. (Collection of abstracts of doctoral dissertations and monographs available in complete form on microfilm. A yearly index is provided.)

Several other lists of titles and abstracts of graduate theses are published from time to time by various colleges and universities.

Books and Periodicals

Library card catalogues provide a source of reference to books and periodicals housed in the local library. However, the following list of books and periodicals is essential to consider in searching for leads prior to the use of the card catalogue:

Alexander, Carter and Burke, Arvid J., *How to Locate Educational Information and Data.* New York: Bureau of Publications, Teachers College, Columbia University, 1950. (This is a guide to guides. It provides an aid to quick utilization of the literature in education.)

Winchell, Constance M. *Guide to Reference Books.* Chicago: American Library Association, 1950, 75th ed. (An excellent guide to all references.)

Bibliographic Index, New York: H. W. Wilson. Quarterly, 1938–date. (This is a bibliography of bibliographies and is kept up to date.)

United States Catalogue. New York: H. W. Wilson, 4th ed., 1928. (Presents an extensive list of books published in America, English books imported by American firms, and English books imported by Canadian firms from 1898 to 1928.)

Cumulative Book Index. New York: H. W. Wilson. Monthly, 1926–date. (Lists nearly all books published in the English language. It is a continuing supplement to the United States Catalogue.)

Education Index. New York: H. W. Wilson. Monthly, 1929–date. (This is a most useful guide to books and periodicals in education

since January, 1929. It lists nearly all literature from significant American, and a few British, sources.)

Poole's Index to Periodical Literature. 1802–1881, 3d. ed. Supplements to 1906. Gloucester, Mass.: Peter Smith, 1938. (This has long been the source for nineteenth-century periodical literature. Literature is reported under topical headings only, not by author.)

Reader's Guide to Periodical Literature. New York: H. W. Wilson. 1900–date. (Covers articles of popular and general nature listed topically and by author.)

International Index to Periodicals. New York: H. W. Wilson. 1920–date. (Covers about 250 journals containing articles on pure science and the humanities. Prior to 1928, it covered the technical and specialized research articles in education.)

Vertical File Index. New York: H. W. Wilson. Monthly, 1932–date. (Lists pamphlets and ephemeral free and inexpensive materials. Provides reference to materials that have small circulation or may be out of print.)

Biographical Directories

Many times it is necessary in research to include biographical facts concerning leaders in education, scholars, scientists, philosophers, institutional presidents, etc. A few of the directories listing educators especially are as follows:

Leaders in Education, A Biographical Directory. Lancaster, Pa.: Science Press, 1932–date.

Who's Who in American Education. New York: The Robert C. Cook Co., 1928–date.

Presidents and Professors in American Colleges and Universities. New York: The Robert C. Cook Co., 1935–date.

Who's Who in America. Chicago: The A. N. Marquis Co., 1899–date.

National Register of Scientific and Technical Personnel. Washington, D.C.: American Psychological Association and the National Science Foundation, 1940–date.

SELECTED REFERENCES

Alexander, Carter, and Burke, Avid J. *How to Locate Educational Information and Data.* New York: Teachers College, Columbia University, 1950. xvix + 441 pp. (Read chaps. II and III.)

Brickman, William W. "Reference Aids in Educational Research." *School and Society* (May 27, 1950), *71*:324–331. (A list of bibliographies on research.)

Carroll, John B. "Neglected Areas in Educational Research," *Phi Delta Kappan, 42*:339–343, May, 1961.

Chambers, M. M. "Selection, Definition, and Delineation of a Doctoral Research Problem," *Phi Delta Kappan, 42*:71–73, November, 1960.

Gage, N. L. (Ed.). *Handbook of Research on Teaching.* Chicago: Rand McNally, 1963. (See Parts III and IV.)

Good, Carter V. *Introduction to Educational Research.* New York: Appleton-Century-Crofts, 1959. (See pp. 40–71.)

Graham, J. "Planning and Selecting the Problem," *National Business Education Quarterly, 24*:5–10, March, 1956. (Brief article listing basic considerations in finding and selecting a problem and the steps in testing tentative problems.)

Hardway, Charles. "Values of and Incentives to Research." *Teachers College Journal* (November, 1949). (A discussion for beginners in research.)

Kerlinger, Fred N. "Practicality and Educational Research," *School Review, 67*:281–291, Autumn, 1959.

Lessenberry, D. D. "Research Is No Better Than the Method Used," *National Business Education Quarterly, 11*:5–6, May, 1943. (There is no pattern for all research workers to follow except insofar as there may be a pattern for any organized thinking.)

McCormick, Thomas C., and Francis, Roy G. *Methods of Research in the Behavioral Sciences.* New York: Harper & Row, 1958, 244 pp. (See pp. 14–23.)

Mouly, George J. *The Science of Educational Research.* New York: American Book, 1963. (See chap. 5.)

Strauss, Samuel. "On Research Ability in Graduate Students," *Journal of Higher Education, 32*:443–448, November, 1961.

Van Dalen, Deobold, and Meyer, William J. *Understanding Educational Research.* New York: McGraw-Hill, 1962. (Refer to chap. 7.)

CHAPTER III

Planning and Organizing the Research Project

Statement of the Problem

Nature and Source of the Data Required
Primary and secondary sources
Availability, authenticity, and adequacy of information

Techniques for Collecting Information

Considerations for Presenting and Interpreting Data
Assumptions and hypotheses
Classifying and presenting data
Considerations for analysis of data

Possible Conclusions and Implications

Major Considerations in Sampling
Purposes of the investigation
Population to be sampled
Sources of the population
Size of the sample

Types of Sampling
Random
Stratified
Area or area-probability
Multistage

Once the general problem area has been decided upon, the next step for the researcher is to plan the actual conduct of his research in such a way that he can efficiently attempt its solution. He must first carefully study the historical development of the problem to acquaint himself with (1) the reasons that it occurred in the first place and why it persists, (2) previous efforts at solution, (3) whether the problem is becoming more or less severe, (4) the principal individuals concerned at various times, and (5) other factors necessary to understand just how and why it now exists in the form it does. A full and complete appreciation of its background is absolutely essential to the development of a theory and an adequate statement of the problem. It may enable the researcher to make a few tentative projections regarding the possible nature of the problem in the future unless it is solved or its effects mitigated in some way. This orientation phase may involve considerable library research and possibly the collection of some direct information.

As soon as the problem has been placed in the proper theoretical context, it can be delimited in a number of ways to bring its scope within the researcher's purposes and abilities. Delimitation consists of making choices from a number of alternatives, such as coverage by geographical area, type of school system, and time period. A considerable amount of contraction can be effected by the exclusion of borderline cases or questions. However, every contraction in scope restricts the conclusions that may be drawn; it either narrows the range of factors available for analysis, or reduces the number of cases or numbers and classes of statistical units to be analyzed. On the other hand, the inclusion of borderline aspects may greatly enrich and improve the study.

Three important factors not directly related to the nature of the problem are important in setting the exact boundaries of the investigation. These were mentioned in Chapter II and are repeated here because they determine whether or not the topic is "researchable." In the first place, there must be a clear understanding of exactly what information is necessary to complete the research successfully; next, the feasibility of ob-

taining it must be kept in mind; and finally, the researcher must be sure that the cost of getting the necessary information and of processing it will be within the financial limits of the project. In many instances, there are also very important time limitations involved.

Once these preliminary steps have been taken, the researcher should define his project as precisely as possible and then develop a research proposal or design which will guide his course of study. This proposal, headed by a precise, descriptive, and brief title, should detail (1) the nature of the problem, (2) the necessary information to be obtained, (3) the general procedures to be followed in acquiring this information, (4) the techniques for presenting and interpreting the data, (5) the expected findings and hypothetical conclusions, and (6) any recommendations which would be based on, or any implications which would arise from, the expected results. In organizing this proposal a scheme should be developed to insure that the answer to every question raised will be based upon adequate data and a satisfactory procedure for treating them. For example, if there are three postulates or major parts to the problem, they should be stated as A, B, and C. Moreover, the data needed, hypotheses, procedures, anticipated findings, conclusions, and recommendations should be stated under the respective headings. Thus there is less chance of important elements in the study being inadvertently overlooked.

At this point in planning the research project, criteria should be selected for identifying what data are to be used, how they are to be collected, how they can be classified most appropriately, and for determining the methods of analysis to be used. If two or more research studies on the same general problem use different standards or criteria, it is possible for them to arrive at radically different, yet equally valid, conclusions.

The researcher should have his proposal approved by the institution or agency through which the research is to be conducted before starting any extensive work on the investigation. Thus, the researcher should consider various methods and techniques in developing his general plan for carrying the study

through to completion. The remainder of this chapter is directed toward an amplification of the basic considerations in the development of a research proposal with respect to the six points presented above. In addition, a section of this chapter is devoted to the "sampling concept" because sampling in some form or another is usually involved in most research work, and many decisions with respect to general sampling procedures *must be made at the research proposal stage.* More specific sampling techniques will be discussed in subsequent chapters as they apply to particular methods of data collection and analysis presented at that time.

STATEMENT OF THE PROBLEM

The statement of the problem for a research proposal should include some indication of the source of the problem and its justification in terms of general value. It should indicate the scope of the problem by stating briefly the kinds of persons, materials, and situations to which it is to apply. The specific investigation proposed should be shown to be part of some general problem of greater magnitude and importance. The following excerpts from Aikenhead's[1] proposal for a doctoral dissertation in education are illustrative.

[1] John Douglas Aikenhead. A typewritten doctoral dissertation proposal submitted to a committee of the faculty of the School of Education, University of Oregon, June 3, 1952. The completed dissertation is on file in the University of Oregon Library: "To Teach, or Not to Teach." (Doctoral dissertation, typewritten.) Eugene, Ore.: University of Oregon, 1954.

PROPOSED STUDY TO CULMINATE IN DISSERTATION FOR D.ED. DEGREE

TITLE: Influences Which Cause Persons to Teach, or Not to Teach, in Western Canada

The broad problem. It may be assumed that the improvement of society might be accelerated by better schools, and because almost all persons pass through the schools, more adequate staffs would be effective. Some schools in western Canada are without teachers, and some are staffed by teachers who hold substandard qualifications. The supply of teachers may be mainly influenced by certification requirements, money spent on schools, and the community attitude. Below the state level, school government covers policy developments, enactment, and execution which may be redivided into these overlapping groups: finance, curriculum, public relations, business administration, and personnel.

Personnel may be further subdivided: children and young people as learners; clerks and maintenance help; and the professional, mainly instructional help, but including principals, supervisors, the medical staff, curriculum, and research specialists, as well as the superintendent who develops and enacts policy for the board of education.

A more limited problem. How can the teacher shortage be best alleviated? This problem has breadth and depth. "How far back must we look to understand the meaning of what is happening now? . . . What is the sequence of key facts? . . . How fast is the pace?"[1]

---

[1]Paul Pigors and Charles Myers. Personnel Administration. New York: McGraw-Hill Book Co., 1947, pp. 39-41.

Western Canada's teacher shortage seems to focus on recruitment, retaining teachers, and persuading

them to return to the classroom. Why do young people in high schools choose to teach, or not to teach; why do some teachers leave the profession, or remain; and why do some return after an asbence from the classroom? These three elements will be subdivided in some detail, as questions, but not so inclusive as the questionnaire for this proposed study.

Importance of the problem. The teacher shortage in the Canadian provinces is 10,860[2] with the situation in the western provinces becoming worse in 1951 and 1952. Figures for the United States show a similar trend for the elementary schools. Estimates by the National Commission on Teacher Education and Professional Standards claim the decade, 1948-1958, will require 1,033,994 new teachers in the elementary school, and that the accumulating shortage may reach 800,000 by 1958.[3] Ray C. Maul, who issues periodical reports on the teacher shortage, claims that 500,000 teachers now in service have inadequate preparation; that the present over-crowding of pupils needs 10,000 teachers right away to solve this problem; and that the elementary school population is expanding enormously.[4]

---

[2] The Canadian Education Association. The Status of The Teaching Profession. Toronto, Canada: The Association, 1948, p. 107.

[3] E. W. Anderson and R. H. Eliassen. "Supply and Demand in Teaching." Review of Educational Research, (June, 1949), 29:182-183.

[4] Ray C. Maul. "A Look at Our Teacher Personnel Needs." American School Board Journal, (April, 1952), 184:29-30.

Up to this point Aikenhead has presented an over-all problem of large magnitude, has limited it to the subdivision of "personnel," and further limited it to "teacher recruitment and retention." He has indicated the need for such a study and has buttressed his remarks with citations to authority. A more specific statement of his problem may be seen in the breakdown of three major issues into questions for which he will seek answers, as follows:

<u>Recruitment.</u> What induces young people to teach?

A. Are impressions and school guidance persuasive?
   1. Are the years 11 to 17 critical?
   2. Do parents and relatives influence strongly?
   3. Do others exert a marked influence?

B. Is money a strong influence?
   1. Do young people consider salary first?
   2. Are they aware of the wages currently paid?
   3. Are supplementary benefits considered?
   4. Does the cost of finishing high school deter?
   5. Is the cost of training a handicap?
   6. Would subsidies help for teacher training?
   7. Do rural conditions discourage young people?

C. Does the school plant attract prospective teachers?
   1. Is the building cheerful, comfortable, well lighted, well equipped, and modern?
   2. Is there adequate space for play and good play equipment?

D. Are other influences persuasive?
   1. Will young people be proud to teach?
   2. Will they enjoy peer and community status?
   3. Will they have personal freedom out of school?
   4. Do they feel marriage may be unduly postponed?
   5. Are they fearful of critical parents and rude statements?
   6. Do they feel teaching offers security?
   7. Did one of their outstanding teachers influence them strongly?

<u>Retaining</u> <u>Teachers.</u> Why do teachers stay with the task?

A. Are living conditions as good as the community average?
   1. Is the house attractively located, with landscaped grounds, cheerful inside and out, roomy, warm, and modern?
   2. Is the rental, or payments, realistic in relation to the teacher's salary?

B. Are the salaries and other benefits depending on money adequate?
   1. Can the teacher enjoy as high a standard of living as other people with similar training?
   2. Does the teacher feel the money invested in training has been well spent?
   3. Is there assurance of sick leave benefits and a reasonable pension?

C. Does the teacher enjoy community and peer status?
   1. Is the school an important institution in the community?
   2. Does the community make an effort to orient new staff?
   3. Does the teacher feel that leaving the classroom for an office position, or the kitchen, would improve status?
   4. Is the teacher's personal freedom restricted?

D. Are working conditions genial?
   1. Is the school plant cheerful?
   2. Is there a room for the teachers to enjoy in private?
   3. Is there an adequate supply of curriculum materials?
   4. Is there time during the school day for planning?
   5. Does the staff have a planned time for social contacts?
   6. Do they work as a group on curriculum and staff problems?

7. Is evaluation considered mainly the teacher's responsibility, with the principal and others as advisors?
8. Does the teacher receive praise occasionally?
9. Are other factors present for building morale?

E. Are there opportunities for recreation?
   1. Do married teachers have a normal social life?

F. Are there opportunities for professional advancement?
   1. Is encouragement given to experiment?
   2. Is there a professional bookshelf in the school?
   3. Is the teacher paid for some days at professional meetings while away from the classroom?

<u>Teachers Returning to the Classroom.</u> What will persuade them?

A. Is the need for money a deciding factor?
   1. How much salary is persuasive?
   2. To what extent are age, marital status, and dependents influential?

B. Do ex-teachers feel it their duty to return to teach?
   1. How many came back for that reason in wartime?
   2. Does the present crisis influence them in the same way?
   3. Is the professional code likely to include such a responsibility?

C. Do some teachers return because they wish to keep up to date?
   1. Is the motive largely for professional competence?
   2. Is the return for a short time just to keep their certificate reinstated?

D. Do some teachers return because they feel a desire for change?
   1. If these persons apply, can they be adjusted to the system?

It is obvious that Aikenhead had given considerable thought to the development of his research proposal and presented it in sufficient detail to give the committee a reasonably full and clear understanding of its nature and scope. This problem was considered worthy of investigation for a doctoral dissertation.

A student should keep certain significant points in mind in preparing his own research proposal. The application of these points may be seen in Aikenhead's statement of his problem. In the first place, the work of analyzing the problem for the purpose of discovering its elements, of differentiating these elements, and of rearranging them into a logical sequence must be regarded as a definite outgrowth of the formulation of the problem. A mere haphazard enumeration of points somehow related to the problem in general should be avoided. All questions raised must be related to the problem. Secondly, each major breakdown, issue, or element should be separated into its subsidiary or secondary elements, and these should be arranged in logical order under the major division. Thirdly, a reformulation of the problem is often necessary and desirable. The various divisions may be too widely diversified, the problem cover too great a scope, or contain multiple problems. The original formulation may not strike at the core of the problem, or it may permit of erroneous interpretations. If reformulation is apparently necessary, it should be done before submitting the proposal for acceptance by an adviser or institutional committee.

While further reference will be made to Aikenhead's proposal for a doctoral dissertation, it might be interesting to a student to see how similar procedures were followed in the development of a proposal for a field study by Needham.[2]

[2] John Needham, "The Type of In-Service Program Which Should Be Provided for Oregon Teachers." A proposal for a field study in the School of Education, Eugene, Ore.: University of Oregon, 1949.

TITLE: The Type of In-Service Program Which Should Be Provided for Oregon Teachers.

There has been a growing recognition of the need for teachers to continue their professional

growth beyond the point provided by original academic training and subsequent classroom experience. It is only logical to assume that in-service training programs should be based on sound educational theory and contribute to real professional growth. Teachers should have the knowledge and maturity to recognize their inadequacies. When in-service programs are organized in accordance with the soundest principles developing from research and in harmony with the felt needs of teachers, a greater measure of professional growth will be the end result.

<u>Statement of the Problem.</u> The purposes of this study are: (1) to discover the opinions of teachers in Oregon schools regarding their competency in selected phases of their work and the types of in-service training they deem most valuable in furthering their competency, and (2) to present the most acceptable modern theory and practice as evidenced by the writings of authorities on in-service training.

<u>Justification of the Problem.</u> Many teachers are too complacent concerning their teaching ability and efforts to maintain continuous professional growth. Such an attitude is not compatible with the most effective teaching in a dynamic world society. Knowledge of teachers' opinions concerning their competency and the effectiveness of various types of in-service training, coupled with the findings from library research, should facilitate the writer's future efforts to promote professional growth as a teacher and administrator.

<u>Limitations of the Study.</u> Research on the problem will include all phases of in-service education but this report will be restricted to more specific mention of the techniques included in the attached check list [omitted from this example]. No attempt

will be made to discover why teachers have their opinions on the value of certain in-service techniques, but simply to report these opinions as such. Finally, no attempt will be made to measure the validity of teachers' opinions regarding their competency. The justification for these limitations is that the study is primarily concerned with determining underlying theory with an acceptance of teachers' opinions as an indication of professional insight. This latter factor must be considered, regardless of the merit of opinions, in the establishment of any completely successful program for the promotion of professional growth.

If this proposal of Needham's were to be submitted for consideration as a doctoral proposal, it could not be so sharply limited. It would have been necessary to establish the validity of the teachers' opinions, and probably to present some sort of an experimental program, or programs, of in-service training to determine the most desirable. This particular study, however, might well be a point of departure for the establishment of experimental programs and other researches of doctoral scope and calibre.

In summary, the introductory section of a research proposal should present a general description of the topic, an amplification of its purposes, a justification of its selection, a precise title, and the specific questions to be answered.

NATURE AND SOURCE OF THE DATA REQUIRED

When the problem has been formally stated, the researcher's next task is to set forth some idea of what direction the investigation should take to attain the desired end. This is somewhat similar to our actions when we lose something. Suppose one morning I am unable to find my keys. I may look for them in a helter-skelter manner, or I may attempt to organize my search. If I do the latter, I will probably attempt to reconstruct my movements during the period I last had my keys. This may lead me to several possible locations where the keys might be.

Finally, one of these will be right and I will have recovered my keys. Starting out to collect data is much the same; it must be done in an organized way and this necessitates a *working hypothesis*[3] as a guide for our actions.

A working hypothesis may be adopted with very little information. There are many possible explanations of the phenomena which the researcher is studying, and he may begin to organize his work around the one which seems most reasonable to him—and which may be no more than a guess. But as he continues his investigation, he finds that his tentative hypothesis or guess needs adjusting. In other words, when the original hypothesis is found wanting, a new one may be substituted, and so on until at last one is adopted which provides the best explanation the researcher can prepare.

Once he has arrived at his working hypothesis, the researcher is ready to consider carefully what data are needed to complete the project. In every project at least some essential information may be determined indirectly. Since measurement tends to be an integral part of scientific research, it is necessary to learn what data may be measured directly and what must be measured indirectly. Direct measurement is most easily applied to elements that can be expressed *quantitatively,* such as numbers of workers, age of machines, and square feet of space used for selling merchandise. Indirect measurement is associated more with *qualitative* elements, such as efficiency, skill, aptitude, temperament, and motives. It is frequently desirable in the research proposal to include, following the specific questions for investigation, not only the kind, but also the amount of data required to answer the questions satisfactorily.

Primary and Secondary Sources

In general there are two sources of data—primary and secondary. Data from primary sources are those that are collected

[3] Working hypotheses should be distinguished from the more formal hypotheses which which we are familiar from logic and principles of statistical inference. Formal hypotheses must be rigorously defined and stated in terms of numerical quantities if they are to be amenable to statistical tests of significance.

for the first time as part of such a project. Data from secondary sources are those that have been collected previously and reported by some individual other than the present researcher.

For example, if one were to attempt to find out which of two diets would result in the greatest weight gain for chickens during the period from incubation until three months of age, a primary source of data would be the original experiment. A researcher obtains a number of hatching eggs certified to be from a specific strain of chickens. He places them in an incubator and the chickens are hatched. He separates the chicks into two groups, one of which is fed diet A and the other is fed diet B. All factors except the diets of the chicks are made as similar as possible. At the end of three months the chickens are weighed to determine which group has gained the more. Under the conditions of the experiment, the researcher can come to certain conclusions with respect to the effect of the two diets. The detailed records made on the spot constitute *primary data* to that researcher.

A second source of the data would be available to others if the researcher wrote up a report of his experiment. In this process he would describe as accurately as he could his procedures, his findings, and the conclusions he reached. Although this would seem, on its face, to be a reliable and authentic source, in the process of writing his report the researcher might unconsciously introduce interpretations that were not based entirely on his work, or he might not have the ability to write his report so that a reader would have the same perceptions that he had.

A third source of the data might be a business or agricultural bulletin in which an account of the research is published. This bulletin would probably present a digested or abstracted account of the problem, the research procedures, findings, and conclusions. The editor of the bulletin, or the writer of the article, in reading and writing up an account of the previously reported research might introduce other misinterpretations due to reading the study from a different point of view and not knowing exactly how the project was conducted except from

what he understood by reading the previous report. In any event, a digest or abstract necessarily omits a great deal of material in order to attain briefness, and some of the flavor of the original study is lost.

A fourth source of data might be a textbook on the raising of chickens which included a summary of this experiment along with many others. This might have been taken from the original report, or it might have been abstracted from the business or agricultural bulletin.

A fifth source of data might be a review of reports, as listed in several textbooks and professional or trade journals, in which the previous report of the experiment has been still further digested.

All except the first are technically classed as *secondary sources.* There is little doubt that the last would not be as reliable as the second. In every repetition of one writer reporting the study as described by another, there is apt to be a loss in reliability. This fact has been attested to time and time again by the simple parlor game, "Secrets," with which nearly everyone is acquainted. The enjoyable part of this game, in which a secret is started by one person and passed from one to another until it is finally told by the last one in the group, is to see how such a process distorts the original account.

Primary sources are usually considered to be the most reliable because the researcher, himself, is more able than anyone else to understand and interpret what he recorded and to know the conditions under which the data were obtained. Secondary sources must be investigated to find out the conditions and the purposes for which the data were originally collected, and the distortions in reporting which might have taken place since their collection and subsequent publications. For practical purposes, however, the second source mentioned above —the researcher's own report of his experiment—may be considered primary since it is the only evidence that would be available to anyone other than the original researcher. If it is impossible to obtain primary data, a basic principle of data collection is to obtain data closest to the primary data.

AVAILABILITY, AUTHENTICITY, AND ADEQUACY OF INFORMATION

After the kinds of information needed have been ascertained, the researcher must determine whether or not they are available to *him*. It is often more difficult to obtain data than it is to determine what data are needed. For example, in a project requiring information about individual salary payments made by a firm to its employees, it may be simple enough to learn that these data are in the firm's accounting records, but it may be quite a difficult task to obtain the information because of the firm's reluctance to make such records available to the researcher.

The researcher must check the authenticity—that is, the accuracy—of the data he collects. In the above example, the researcher may not be able to get access to the firm's records, but may obtain reports from employees through interview or questionnaire procedures. Since some employees misrepresent their earnings when questioned, the data they supply may be invalid for research purposes.

In referring to the previous example of a research proposal by Aikenhead, the following statement with respect to the determination of data was submitted.

<u>Sources of Data.</u> Alberta, British Columbia, Manitoba, and Saskatchewan, respectively, have: 5003, 6039, 7001, and 4023 teachers regularly in the classroom to total 22,066.* The aggregate of students finishing high school would be about 20,000, with a conservative estimate of ex-teachers placed at 30,000.† An attempt will be made to get a cross section of these by sending a questionnaire in equal numbers to these five groups:

1. Students in the last two years of high school who are not interested in training to teach.
2. Students in the last two years of high school who are serious in their intention to teach.

3. Teachers, now teaching, with less than ten years of experience.
4. Licensed teachers, who have returned to teaching after an absence of at least eleven months.
5. Former teachers, below 60 years of age, who had at least three months professional training, and who are not now teaching.

Assumptions: (1) The persons receiving the questionnaire are capable and willing to formulate and express their opinions. (2) These persons will be forthright in most of their answers. (3) The expressed opinions may be based on the true motives and thereby a reasonably accurate portrayal may result.

* Canadian Education Assocation, Newsletter (May 9, 1949). Toronto, Canada: The Association, p. 1.

† Forty-Sixth Annual Report of the Department of Education: Government of Alberta. Edmonton, Alberta: The Queen's Printer, 1951, p. 126.

In this proposal, Aikenhead had made an effort to determine the approximate number of students, teachers, and ex-teachers who could be considered as the population from which he would draw a representative sample. Since he was to obtain his information from a mailed questionnaire, he had to make some assumptions with respect to the responses he could expect, but there could be no more reliable and authentic source for his study than asking the questions of the individuals themselves. The validity of the responses, however, could only be assumed in terms of his proficiency in constructing his questionnaire to elicit true statements. Techniques in this respect are considered in Chapter VI.

The adequacy of information refers to whether it is sufficient in amount to serve as a basis for drawing valid conclusions. In many studies, especially those in which formal hypotheses are to be subjected to tests of statistical significance, it may be difficult for a researcher to obtain a sufficient amount of data

to make good analyses possible. This point is discussed in more detail later in this chapter with respect to sampling considerations, and also in Chapter VIII with respect to the development of experimental designs for research.

TECHNIQUES FOR COLLECTING INFORMATION

Direct collection of data, as opposed to the use of data already collected by others and made available in published form, necessitates the use of carefully planned guides. The word "schedule" is used to describe any systematic set of questions with space for the researcher to record entries in an orderly way. It is most often used in studies that rely on quantitative methods of analysis, but it is also adapted for the recording of qualitative information, because it assures that the basic points of the inquiry will be covered.

Preparing a good schedule requires a great deal of effort, and temporary drafts should be tried out in pilot studies until a form is developed that is reasonably satisfactory. The researcher is then ready to conduct his investigation using any technique for collecting information which is mentioned in the following paragraph. Details of schedule construction for specific purposes are presented in Chapters IV, V, VI.

The procedures for obtaining primary data are described in the literature under various "methods" of research including historical, normative-survey, descriptive, experimental, genetic, and case-study; but these are in general only modifications of three basic procedures—*observations, interviews,* and *correspondence.* While specific techniques for each of these will be discussed more fully in subsequent chapters, at this point I shall present a few basic considerations which may aid the researcher in understanding the relative merits of each procedure. Naturally the desirability and validity of each depends upon the information required, and the researcher should select the technique or techniques best suited to his purpose in the early stages of planning his study.

Usually, personal observation is the best method for obtaining primary data as the researcher knows what he has in mind

and can make the appropriate interpretations. In addition, there are many topics, especially those dealing with aspects of human behavior, which can be investigated satisfactorily in no other way. Observations may be made under actual or real conditions, or under simulated and arbitrarily created conditions such as "business games." These are often "played" by using electronic data equipment. Observations may also be made by mechanical instruments which record how individuals behave in particular situations; such an instrument is the device attached to a TV set which records the channel to which the set is tuned at different hours. Observation techniques may also record how physical phenomena react in natural states or under other various conditions. In many education studies which utilize techniques more commonly adopted by researchers in sociology and psychology, the investigator becomes one of a group of persons whom he observes—this is participant observation. He may participate actively or merely as an onlooker. In any of the above cases a schedule must be prepared so that the observations can be carefully recorded for later interpretation.

The interview method necessitates asking someone else for factual information. In other words, the researcher obtains as best he can observations or interpretations that have been made by another person. Of course, this process allows for ambiguities and misinterpretations; the respondent may misunderstand the researcher's questions and the researcher may misinterpret the answers of the respondent. Fortunately, it is a face-to-face situation in which certain confusions will be immediately apparent and can be quickly corrected. The interview method also provides an opportunity for the exchange of thoughts and for the observation of facial, physical, and vocal reactions not obtainable by some other methods. The telephone conversation is a modified form of this technique, but lacks the face-to-face advantages of the personal interview.

The use of correspondence in the form of personal letters or mailed questionnaires permits the researcher to collect information from sources which might otherwise be inaccessible to him. Its greatest advantage is that it permits the collection of data

from large numbers of individuals at a relatively low cost and in a period of time which would not be possible by other means. The major disadvantage of this method is that further ambiguities may be introduced because of misinterpretation of the questions submitted and the answers given, the incompleteness of returns, and the careless completion of those questionnaires which are returned. Techniques for reducing ambiguity are presented in Chapter VI.

When professional journal articles, books, manuscripts, and other documentary materials are used for data collection, no standard schedule can be used, but the researcher should consider carefully certain background information about each document used in order to evaluate its reliability and appropriateness. When preparing a research proposal, the researcher probably can only scan most of the documentary material he hopes will be useful later on in his investigation. What he should do at this point is to build up a bibliography, typically in the form of a card file, which gives a complete description of the book, bulletin, or other written source, and indicates where and how it may be obtained when it is needed, why each entry is included, and what specific pages contain information which may later be useful. The reader is referred to Appendix A for a recommended form for bibliographic notations.

CONSIDERATIONS FOR PRESENTING AND INTERPRETING DATA

The purpose of collecting information for a research project is to use it to draw conclusions and make generalizations which will aid in solving the problem which is being studied. It may be that the information so clearly points to a particular conclusion, generalization, or solution that no aid is required. In complex situations, which is what we are concerned with in this text, such is not the case. The information collected directly or by others must be rearranged; most of it must be condensed and then studied carefully to uncover possible relationships hidden in the mass of detail which constitutes the raw information. Skeleton tables and charts must be devised to serve as a

basis for considering the general methods of analysis and as guides to the collection and recording of data. Before deciding how he will present and analyze his data, the researcher should consider his assumptions and hypotheses.

Assumptions and Hypotheses

Frequently these two terms have been considered to be approximately synonymous; they are similar in that both are propositions accepted tentatively to facilitate the solution of a problem. However, as regards research, a distinction should be made between them. An *assumption* is accepted without thought of immediate proof. It is a proposition for which no information can be made available within the practical limits of the study. Some assumptions are *axiomatic* in that they are propositions that virtually every reasonable person is ready to adopt but which cannot be proven. People assume that nature is uniform, that laws which have applied in the past and are applying at the present time will apply in the future, and that there will always be a "tomorrow"; these are axiomatic assumptions. The type of assumption most commonly stated explicitly is one that is *limiting* in its nature and serves to hold the size or scope of an investigation within its prescribed boundaries. A common one is that man's economic behavior is rational, that he acts to maximize his economic well-being. This particular assumption has been dignified with the title, "the economic man." Another common assumption is that man knows the details of his environment. Assumptions are usually made when the argument rests on *a priori* reasoning, but they may also be made on the basis of present knowledge, or research, which is as yet incomplete. Assumptions of the latter kind are used to avoid the prolonged delay that would result if the propositions were actually to be tested and verified. Accordingly, specific qualifications must be made in the conclusions of research in which these assumptions are made. In any case, assumptions ought to be clearly stated.

In contrast to assumptions, *hypotheses* are propositions to be tested and are the very subject of the investigation. A hy-

pothesis may be defined as a tentative proposition, stated as a generalization, which is to be tested from a sample of data to be collected in a research project. We have already explained the use of "working hypotheses" in connection with the identification of data to answer specific questions in the research problem. At this point, we must select from these "working hypotheses" the ones that appear to be most reasonable and promising and state them in terms suitable for testing their significance. Hypotheses that are to be subjected to statistical tests must be rigorously defined and stated in terms of numerical quantities. That is, we cannot state as a hypothesis, "A is greater than B," since we have no exact numerical value that represents "greater than." Our working hypothesis could have been so stated, but now we must be more precise and state, "A is equal to B," or "A minus B equals zero." Either of these statements, commonly referred to as the *null hypothesis,* can be tested statistically. If we can reject this hypothesis, we are then in the position of saying that one is greater than the other, although we have not as yet determined "how much" greater. Further details for defining and testing hypotheses are included in Chapter VIII.

Classifying and Presenting Data

To facilitate the analysis of large amounts of data, it is necessary to classify the data in some manner. A general rule of classification is to place together the things that possess the greatest number of similar attributes. While data can be classified according to many different attributes, the following five classes cover most situations: *general aspects*—appearance, size, form, age; *components*—various parts, members, elements; *form or structure*—the underlying skeletal framework of the data; *dynamics*—the forces or systems operating within the data; and *progressive stages*—the developmental growth or trends in the data.

When classifying data, the researcher should be constantly aware of apparently large differences, or similarities, among the various categories for his object is to analyze and arrange

the data so finer and finer distinctions can be drawn. The very essence of science is proper classification. It is an attempt of the mind to rationalize things. It is the only way to discover the proper order of things or to show the actual relationships among various kinds of things.

After determining the proper classification of the data, the researcher should edit them, retaining only those that have a clear influence on the problem at hand. These data should be grouped according to their consequences, or combined to produce a common result, and described precisely according to their character and extent. Their subsequent analyses are facilitated by use of various tables, charts, graphs, and statistical calculations. When using any of these, the researcher must determine what facts should be emphasized, how they can be made most evident, and what manner of presentation, whether typewritten or printed report, is best suited to provide the clarity and precision necessary in a research paper.

The simplest method of presenting quantitative research data is to list in one column of a table different qualities or values of data with the frequency of occurrence for each in a corresponding second column. A more complex presentation involving multiple columns is required if the data are extensive. However, these tabulations should not be in a haphazard form, but should be so ordered that the overall characteristics are revealed. The qualities in one column should be mutually exclusive since overlapping values can only lead to confusion. When a fairly large number of items are presented, they should be grouped in intervals that are not so small that the advantages of summarization are lost, nor so large that important characteristics become hidden. Tables should possess unity—in general, the data each presents should be of one type and character. One table, for example, may contain the complete gross data on which the study is based, while another may summarize these data, bringing out their significant features.

The data embodied in a table can often be communicated better by a chart or other graphical form. Pictographs may be used when they serve a real purpose, such as clarifying facts, or

indicating comparisons and relationships. For example, in describing the relative population of states in the United States, sometimes a much more meaningful perception of population distribution may be obtained by constructing a map with the size of the states in direct proportion to the size of the United States as the population of each state is to the total population of the United States. While this procedure distorts the map from the typical Mercator Projection, it does reveal more readily the relative population of the states (see Figure 2). In planning the organization of a research project, the researcher should develop skeletal tables and charts to indicate the types of presentation and analysis he expects to use. Details for the presentation of tables, charts, and graphs are given in Appendix A.

Considerations for Analysis of Data

Probably one of the more fruitful methods of analysis involves contrasting and correlating the size of the items. Many research studies require ranking, making of comparisons, and noting differences among data. Ranking merely involves determining the relative position which any particular item occupies in a group of data of the same kind. The direction of rank may be either upward or downward, with the first measure in either direction being called the first rank, the second measure being called the second rank, and so on. The significance of rank depends on the number of cases in a group. To make ranks meaningful when different groups are composed of varying numbers of cases, the percentile rank is used to provide a common comparative basis for interpretation.

Comparisons are usually determined by any of a number of mathematical procedures to yield indexes of relationship or association—usually by means of correlation coefficients. One primary object of research is to ascertain relationships, if any, among different variables. The relationship most often sought is that of cause and effect. However, the existence of such relationship is difficult to prove; merely because one event happens to follow another in point of time is not proof that the first caused the second. For example, although interest rates gen-

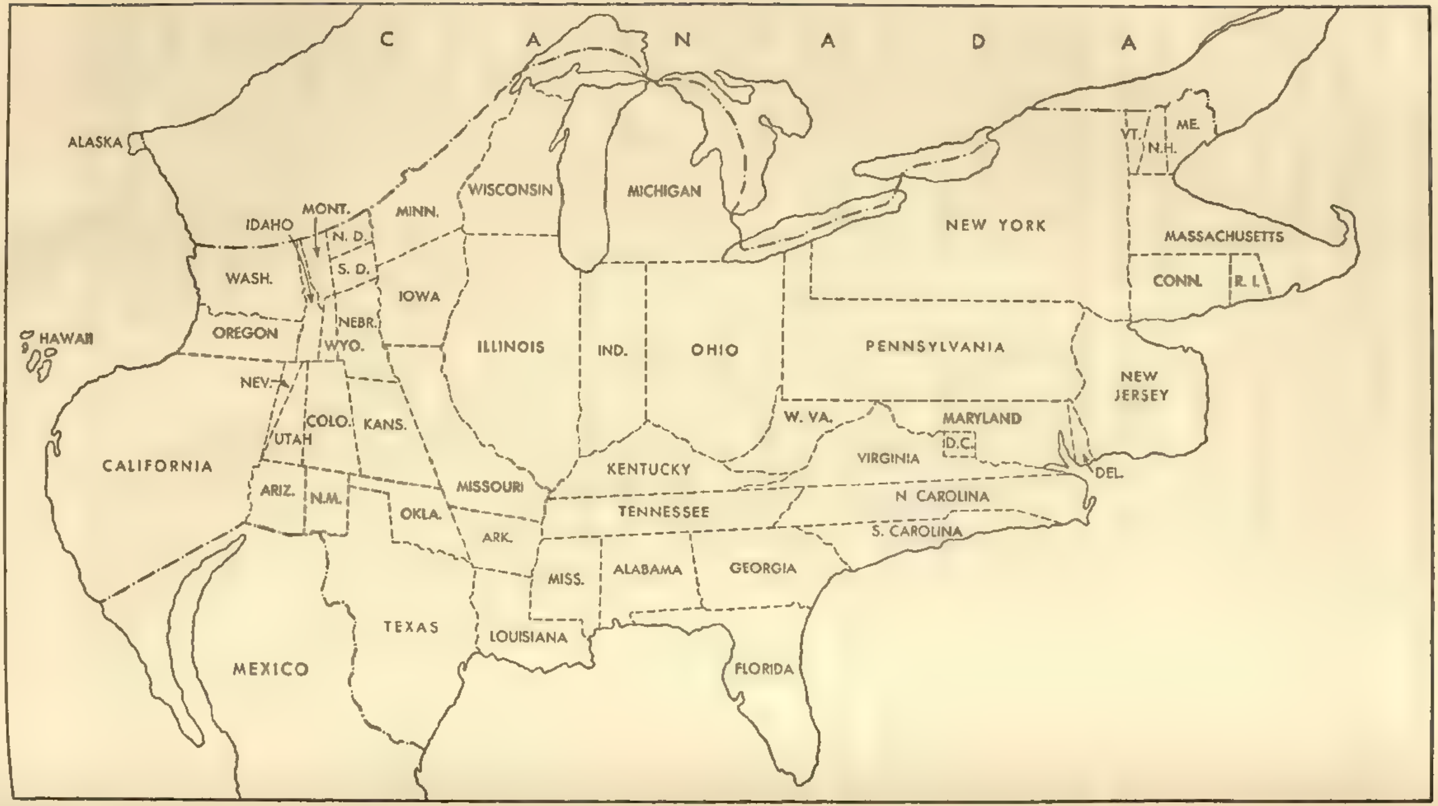


FIGURE 2. Size of States in Relative Proportion to Their Population.

erally rise in the latter part of the prosperity phase of the business cycle, this does not necessarily mean that rising interest rates cause the downturn which may follow. Faulty conclusions about cause-and-effect relationships can be reduced by careful definition and study of the relevant variables for such an investigation may reveal that apparent causal factors are actually the effect of a more basic factor that caused both them and the final effect.

Differences among data are usually treated by various statistical tests to determine whether they are real or whether they could have occurred merely by chance in the sampling process. Another basic object of research, then, is to determine whether or not there are meaningful differences among various subgroups of a population.

Whatever statistical calculations may be needed, they should be described as carefully and precisely as possible at the time the researcher is developing his research proposal. The type of data needed and the method of their collection are dependent to a great extent on the way in which they are to be analyzed. Many novices at research have collected an extensive amount of data only to find out to their sorrow that it was not collected in a manner which would permit satisfactory analysis. While this point will be brought out more specifically in the discussion of the design of experiments in Chapter VIII, the following questions should be of use in developing a tentative plan for analysis:

1. How should the data be grouped?
2. Are the data continuous or discrete in nature?
3. Are the data too finely classified, or subdivided too many times?
4. Can several factors be eliminated so that the reader can study a distribution on the basis of a few factors (or a single factor)?
5. Should some of the data be expressed as percentages of others?
6. What degree of relationship exists between different distributions or classes of data?
7. What differences exist among groups of data?
8. Are the magnitudes of difference great enough to conclude that there was little likelihood of their occurrence due to chance alone?

9. Have the data come from a population in which the characteristics to be analyzed are normally distributed?
10. If a sample is used, has it been selected in such a manner that statistical methods requiring a normal distribution and randomization can be used?

POSSIBLE CONCLUSIONS AND IMPLICATIONS

The final step in the planning and organizing of the research project is to list the types of conclusions that *might* reasonably be expected. While it is impossible, of course, to indicate actual conclusions prior to carrying out the research, it should be possible to indicate some of the results that might, or could possibly grow out of the study. Similarly, each hypothetical conclusion should be considered for its possible implications, as the value of the study depends to some extent upon these implications. If it should be found that none of the hypothetical conclusions has implications of value, there might be some question about the desirability of going ahead with the research. On the other hand, if any or all of the hypothetical conclusions would have valuable implications, the researcher could be assured that the problem would be worthwhile regardless of "how the chips might fall."

MAJOR CONSIDERATIONS IN SAMPLING

In every research undertaking, the researcher must decide whether the entire field, or only a part of it, is to be covered. Frequently we draw conclusions or make generalizations about a whole based on an examination of some part of the whole. The process of using a part as a basis for an estimate of the whole is known as sampling. Technically, every research study follows a sampling procedure in that a particular study covering the entire field at the moment is only a sample of what has been and what is yet to come. Since most research studies are conducted to aid in decisions which will affect future activity, it is necessary to consider the past and the future as being the whole, of which the field at present is only a part, or a sample. Just as one or two classes of fifth-grade pupils may be studied to

determine what could be expected of many other classes of similar nature during any particular time, an attempt to cover the entire field of fifth-grade pupils at the present time would only yield estimates of what could have been expected of similar fifth-grade pupils in years past or what might be expected of fifth-grade pupils in the years to come. Since nearly all studies are made for the purpose of carrying out some activity in the future, on the assumption that the types of conditions existing at the present time will continue to exist for some time to come, it is necessary to consider the past and the future as being the whole.

Practically all research studies in education are of a type in which information and data are obtained from relatively small groups of individuals or from small samples of products. Thus, since their purpose is usually to learn something about larger groups of individuals, products, or "populations," they may be considered sampling studies. In any such study the researcher should attempt to obtain a sample that is truly representative of a larger group or population. Lindquist has pointed out some of the difficulties in obtaining a truly representative sample; these difficulties indicate the need for careful consideration in determining the nature and extent of the sample:

Because the individuals comprising any of these populations differ from one another, and because chance or uncontrolled influences play some part in determining which of these differing individuals are to constitute the sample used, any single fact obtained from a sample (such as a mean, median, percentile, standard deviation, etc.) is almost certain to differ by some amount from the corresponding fact for the whole population. Such "obtained" facts, therefore, may never be accepted at their face value as exactly descriptive of the population involved, but must always be considered only approximations to, or as only *estimates* of, the corresponding "true" facts. So that any such obtained fact may be properly interpreted, then, we need to know how "good" an estimate it is of the corresponding fact for the whole population; that is, we need to have some description of the dependability or *reliability* of the estimate and must qualify accordingly any generalization based upon it. Such descriptions of reliability are extremely important, since without

them we might attribute real significance to facts that are of only accidental origin or read important meanings into mere coincidences.[4]

Any decision regarding sampling must, necessarily, be based upon the established purposes of the investigation, a precise description of the population to be investigated, and the sources of the population from which the sampling units are selected. Once these have been ascertained, the type of sampling procedure to use and the size of the sample to be drawn must be determined.

Purposes of the Investigation

In an earlier section, we discussed the purposes of the investigation. These purposes must be established precisely to learn the type and extent of sampling necessary for a satisfactory solution of the problem. For example, if the purpose of a study is to investigate a problem that is restricted to a particular class in a particular school, the selection of the sample and the conclusions from the investigation must be restricted to that particular group. However, if the purpose of the study is to determine some procedure that may be applied to groups of schools, a community, a city, a state, or some larger "population," then the sample must be determined so that the conclusions of the study can be generalized to the broader population as defined.

Population to Be Sampled

The definition of a population is fundamental and has too frequently been given inadequate attention in research studies. A study does not have to apply to the whole human race, or to the whole population of the United States, to be scientifically valuable. It may be limited to a single school in a single city to which some friendly administrator has given the researcher access. However, the "population" considered must be defined precisely and the significance of the findings of the research study must be limited to whatever population is sampled.

[4] Everet F. Lindquist, *A First Course in Statistics*. Boston: Houghton Mifflin, 1942, p. 102.

The term *population* is often referred to as the *universe;* and, as related to the sample, is sometimes called the *parent, parent population, parent universe,* or *aggregate.* The various units comprising the universe or population are sometimes referred to as *units, elementary units, natural units, ultimate units, elements, members, individuals, items, measures,* or *specimens.* At times intermediate groupings are made composed of groups of units; these are referred to as *chunks, slices, segments, clusters,* or *strata* of the population.

Sources of the Population

In determining the sources of the population from which the sampling units are to be selected, the researcher must relate his description of the population to the purposes of his investigation and establish the boundaries, or *frame,* of the population according to the characteristics of the sample units to be included and their scope. This frame should define the categories of materials or individuals to be covered in the investigation and define the geographical scope and the time elements within which it is to be carried out. The frame is also referred to as the *field of inquiry.* It may consist of available descriptions in the form of maps of geographical areas, lists of individuals, directories of schools, files or records or systems of accounts, photographs, or other descriptive or statistical data in a variety of forms from which sampling units may be selected.

The frame, or field of inquiry, may have several subdivisions made up of *domains of study* or *blocks of domains.* Any subdivision about which the inquiry is planned to provide usable information may be termed a domain of study. Each domain may be expected to provide reliable information about some aspect of the investigation. Blocks of domains consist of two or more domains for the purposes of making comparisons or for evaluating differences among them.

The sampling units selected within the frame may be the same as the elementary units or they may be groups of such units possessing all the characteristics required. For example, a systematic pattern of elementary units might constitute a

sample-unit; households or families are frequently better sample-units for surveys than are individual persons. It is important to describe carefully the qualitative and quantitative characteristics of these sample units, because they will form the basis for all subsequent analysis of the problem under investigation.

Difficult problems of sampling sometimes arise when it is not possible to identify completely the population to be sampled—when the frame is not perfect. A frame may be imperfect because some of the elementary units fail to match the units in the population; or because some of the units are missing, included more than once, or even nonexistent; or because essential descriptive data about the elementary units are unavailable or inaccurate; or because a frame selected at one time becomes out of date before the investigation is concluded. In many cases the researcher will have to correct or modify the frame after its initial construction; this is especially so at the various stages of multistage sampling (see p. 79).

It is difficult to define a frame which will guarantee a complete description of the population, although many researchers have used such sources as census data, telephone directories, automobile owner lists, utility subscriptions, voting registrations, city directories, school directories, court or police blotter records, and the like.

These sources, however, serve only as partial lists and frequently do not define the population that is being studied. Due to the mobility of the population, census data are usually inaccurate by the time they are available to a researcher. Telephone directories, in addition to being out of date soon after their publication, often have a disproportionate number of higher-income families and do not list persons in the lower income brackets who have no telephones or those in the higher income brackets who have "unlisted" numbers. Automobile owner lists represent a select population, varying with the age and cost of automobiles. Court records provide identification of only those individuals in the population who were apprehended and do not include all those who have actually com-

mitted offenses but have not been caught. The determination of the frame is basic to any procedure of sampling, since the sample can only be representative of the frame itself.

Size of the Sample

The determination of the size of a sample is one of the more difficult problems facing a researcher, but it is also one that is uppermost in the minds of most beginners in research. Many novices attempt to start their research project with the determination of the amount of "work" they will have to do instead of planning to solve a problem regardless of the work entailed. Admittedly, it is often necessary to be "practical" and attempt to get the best answers possible with the least amount of effort.

It is easy to say that the sample should be adequate and representative of the population, but "adequate" and "representative" are ambiguous terms. It would be preferable to say that the proper size of sample depends partly upon the purposes of the study and partly on the nature of the population studied. Ordinarily, the sampling error—or, in general terms, the amount of variation that may be attributed to chance elements—varies inversely as the square root of the number of cases drawn in the sample. That is, we would expect to find about twice as much sampling error in a sample of a given size as we would in one four times as large.

In general, the more variation that exists in the population with respect to a characteristic being investigated, the larger the sample ought to be. If information is available about this characteristic from previous studies of the population, various sampling error formulas[5] may be applied from which an indication of the minimum size of the sample may be obtained. For example, a very simple test of the significance of differences between the mean values of two sets of data based on equal size samples might involve a statistical test using the formula,

[5] Sampling error formulas differ considerably according to the nature of the data and the methods of analysis. These should be considered carefully at the time the specific statistical procedure for analysis is being determined.

$$t = \frac{M_a - M_b}{\sqrt{\frac{\sigma_a^2}{n_a} + \frac{\sigma_b^2}{n_b}}}$$

From previous studies it might be found that the means of *A* and *B* were 128 and 120, respectively, and that the standard deviations of each group were approximately 20. If the *t*-ratio (see Appendix B) had to equal or exceed 2 in order to consider the differences between the means as being greater than could be attributed to chance fluctuations in sampling, the number of cases in each group necessary to obtain this *t*-ratio could be determined by substituting the "known" values in the formula and solving for the number of cases (n_a and n_b) needed for the sample. Thus,

$$2 = \frac{128 - 120}{\sqrt{\frac{20^2}{n_a} + \frac{20^2}{n_b}}}$$

and by simple algebraic processes the minimum number of cases in each sample should be 50. To be on the safe side, however, a researcher should draw a sample somewhat larger than the minimum indicated by this procedure since data collected in his study might differ from previous studies. Without quite accurate prior knowledge of the size of means and standard deviations to be obtained, there is no statistical procedure for determining the size of a sample to use.

In homogeneous data there is less tendency for the sample to differ from the population than there is for heterogeneous data. Consequently, the number of items drawn from heterogeneous data should be larger than if the data were homogeneous. For example: One could logically expect that within a box of 50 apples picked at random from a "Red Delicious" tree, the apples would have many characteristics in common (homogeneous data), and within a box of 50 apples picked at random from a "Yellow Transparent" tree, one would also find common characteristics, but that the apples in the latter box would be likely to differ in general from those in the former box. It would take

a considerably larger sampling than 100 apples to detect the same characteristics for each type of apple if the 100 apples were picked at random from a large number of apple trees of many kinds (heterogeneous data).

From a logical point of view, the size a sample should be depends upon the extent to which the individuals are representative of the population to be studied, the inclusiveness of the sample, the types of groups involved, the number of categories of data required, and the method of analysis of the data. It is absolutely essential that the size of the total sample be large enough to permit valid analyses of subsamples used in the smallest breakdown of data to be made. In order that the smallest breakdown of information be representative, there must be a fairly large number of cases in the subsample (25 to 30 would be the minimum as a rough "rule of thumb") instead of only two or three. Thus, this minimum number for each small breakdown could be multiplied by the number of subcategories of breakdowns, and this product could be multiplied by the number of categories, and so on, for an estimate of the total sample. However, this process assumes that a random selection of cases in each subcategory will be uniform throughout the population —an assumption that does not hold for many human characteristics of populations. Probably, the researcher should plan for a size of sample much larger than this process would indicate to insure a satisfactory number of cases in the smallest categories.

TYPES OF SAMPLING

Too often neophytes in research are governed by the factors of expediency and administrative convenience and draw samples from that part of the population that is readily available to them. Thus, they are likely to obtain a *biased* sample, or one whose measures differ systematically from the corresponding "true" measures of the population. To avoid, or at least to minimize, the possibility of biases, various methods or types of sampling have been developed. These may be classified loosely into two categories; *probability samples* (random, stratified, area, and systematic), all of which are so constructed that statisti-

cal theory may be applied to provide valid estimates of the population, and *judgment samples* (purposive and quota), whose results depend largely upon qualities of personal judgment that cannot be isolated and measured, but to which statistical theory may sometimes be applied if certain assumptions are made.

Random Sampling

It is difficult to define in simple terms exactly what is meant by random sampling as a technique with practical applications to many research studies. Strictly speaking, a random sample is one in which every element in the population has a chance of inclusion in the sample equal to that of every other element. A less rigorous, but more practical, definition is that it is a procedure by which the elements are drawn in such a way that there is no reason to believe that bias will result. If the elements are chosen so that no one element is favored over any other element, we approximate the meaning fairly closely. Several techniques can be used for drawing random samples, among which are the following:

Lottery Method. A list is made of all the cases or elements in the frame and numbers are assigned to each case consecutively. Such numbers are then written on identical slips of paper (or chips, or "Bingo" discs), placed in a receptacle, and mixed thoroughly. Then the number of slips required for the sample are drawn from the receptacle.

Technically speaking, every case does *not* have an equal chance of being selected if the slips are withdrawn one at a time because, as soon as one is drawn, the chances are improved for each of the remaining in the second drawing, due to the reduction of the number of slips left in the receptacle. However, for practical purposes, this method is considered to be unbiased and the cases may be assumed to have been drawn at "random." If it were possible to draw the exact number of slips desired in a single draw, there would be no reduction of cases and the procedure would be strictly random.

The lottery method may be varied by use of a roulette wheel or by flipping a coin. These variations are especially well

adapted to the selection of small samples or to the random assignment of "treatments" to intact groups of cases. If two experimental procedures or "treatments" are to be assigned at random to two intact groups of cases, a coin may be flipped upon a predetermined code—"heads" would give one treatment to one group while "tails" would give it to the other group.

Random Numbers. All cases in the population are numbered serially so that each number has the same number of digits in it: e.g., 01, 02, . . . 27, 28, etc.). A table of random numbers[6] is used to determine the cases selected for the sample. It is so constructed that any digit from 0 to 9 has an equal chance to appear in any given position in the table. When using a table of random numbers the researcher selects any point in the table and reads consecutive numbers in any direction (horizontally, vertically, diagonally). The starting point should be determined *prior* to looking at any number in the table to avoid any suspicion of bias. The numbers read indicate the assigned number of each of the cases to be selected for the sample. The following represents a small section taken from a table of random numbers.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	53	74	23	99	67	61	32	28	69	84	94	62	67	86	24
2	63	32	06	86	54	99	00	65	26	94	02	82	90	23	07
3	35	30	68	21	46	06	72	17	10	94	25	21	31	74	96
4	63	43	36	92	69	65	51	18	37	88	61	38	44	12	45
5	98	25	37	55	26	01	91	82	81	46	74	71	12	94	97
6	02	63	31	17	69	71	50	80	39	56	38	15	40	11	48
7	64	55	22	21	82	48	22	28	06	00	61	64	13	54	91
8	85	07	26	13	89	01	10	07	82	04	59	63	69	36	03
9	58	54	16	24	15	51	54	44	82	00	62	61	65	18	69

Two examples of the use of such tables follow. In the first example, assume that 8 individuals are to be selected at random from a list of 100, numbered consecutively from 00 to 99. We

[6] Tables of random numbers are usually included in most textbooks on statistical methodology.

could start with column 5 of the table of random numbers and select the first eight numbers appearing (67, 54, 46, 69, 26, 82, 89, 15). Since 69 appeared twice in this column, it was excluded the second time. Any column, row, or diagonal could have been chosen for this purpose as well as the one illustrated.

In the second example, assume that from a population of 20 individuals, three groups (A, B, and C) are to be formed with 6 individuals in each. We could number the individuals consecutively from 00 to 19. We could decide to exclude the first two numbers drawn, assign the next six to group A, the following six to group B, and assign the remainder to group C without further selection from the table. We could decide to start reading our numbers horizontally beginning with the second row of the table, and continue with subsequent rows until the first 14 had been selected. Thus, the first two (06 and 00) are excluded; the next six (02, 07, 17, 10, 18, 12) are assigned to group A; the next six (01, 15, 11, 13, 04, 03) are assigned to group B; and the remainder (05, 08, 09, 14, 16, 19) will automatically be assigned to group C.

Stratified Sampling

Stratified sampling is the procedure of dividing the population into subpopulations or domains of study, called *strata,* and then selecting a sample within each. A predetermined number of samples is drawn *at random* from each stratum, each domain, or subsample, which is treated as if it were itself a population for purposes of random sampling.

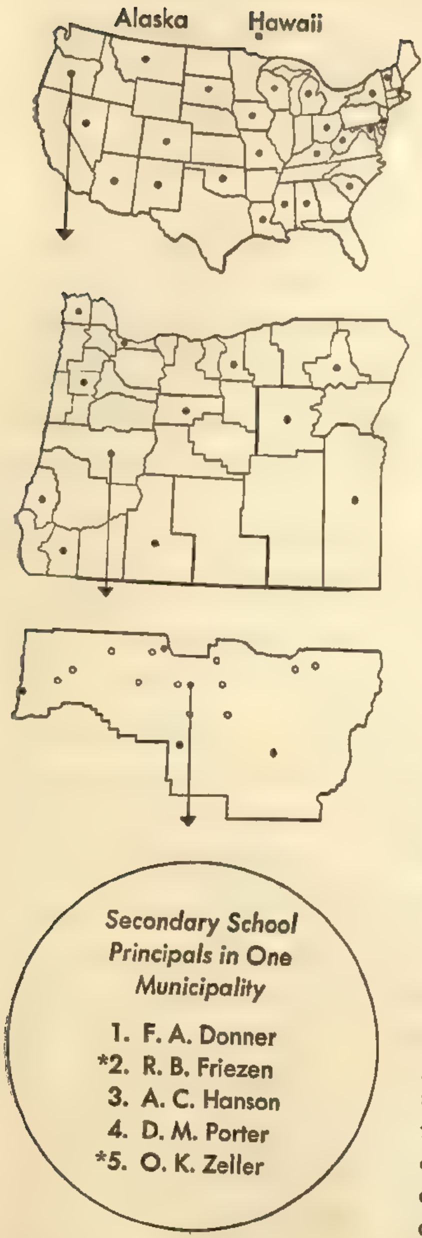
Stratified sampling is used whenever it is necessary to use different sampling methods with different strata of the population, or when there are systematically different characteristics for subgroups within the population. It is also useful when it is difficult to identify and sample individual items at random, but easier by groups of families. However, this sort of stratified sampling, called "cluster sampling," should not be used unless these strata or clusters form identifiable domains of the study, and are not selected merely for the sake of convenience. Another sort, called "proportionate sampling," is based on the

selection of a number of individuals from each stratum which is the same proportion of the total sample as the total number of individuals in the stratum is to the total number in the population.

Area Sampling

Another sampling method, somewhat similar to the cluster and proportionate types of sampling above, is that of *area* or *area-probability* sampling. While it may be applied in many types of investigation, it is especially appropriate when it is desirable to obtain samples representative of various geographic areas. In this process it is not necessary to make a complete list of all the individuals or elements within the entire geographic area or frame.

In a sampling of secondary school principals throughout the area of the United States, for example, the conditions of random sampling of elementary units (principals) can be attained either directly or indirectly. If we were to draw these elementary units directly from the undistributed population of the United States, they would likely be scattered over vast areas with considerable distance between individual units, and the research costs in terms of manpower and travel expenses would make such an investigation impractical for many researchers. On the other hand, if we drew the elementary units selectively, we could reduce our efforts by studying clusters of them. We could divide the geographic area of the United States into *sample-areas* (states, for example) and draw a random sample. The sample-areas drawn would then be divided up into *block-areas* (counties). We could then draw a random sample of the block-areas, divide these clusters into *segments* (municipalities), and draw a random sample of such segments. If a complete enumeration is then made of all the elementary units (principals) in each segment, we could draw a random sample of such principals, and thus fulfill the basic condition of random sampling that all the elementary units in the entire population have had the same chance of inclusion. A schematic diagram of such an area sampling procedure is given in Figure 3.



In making a sample survey, the *population* was defined as all the secondary school principals in the United States.

We divided the population into many smaller *sample areas* consisting of one or more States. The 25 dotted sample areas were selected at random to be in our sample.

Each of the selected sample areas were then divided into *block areas* consisting of one or more Counties. The 12 dotted block areas were selected at random to be in our sample.

Each of the selected block areas were then divided into *segments* consisting of Municipalities. The 5 dotted segments were selected at random to be in our sample.

Each of the selected segments were then divided into *elementary sample units* consisting of one or more secondary school principals. All these units were listed after which one or more of them were selected at random (as indicated by the asterisks) for the collection of the survey data.

FIGURE 3. Area Sampling System for a Nation-Wide Survey of Secondary School Principals.

Multistage Sampling

The area sampling procedure, viewed abstractly, may be considered as *multistage sampling, subsampling,* or the *successive sampling of samples.* The population is regarded as made up of a number of first-stage units, each of which is made up of second-stage units, and so on. A frame is required by which the first-stage units may be defined and selected. Second-stage frames are only required for those units selected from the first-stage units, and so on. The devices of stratification and proportionate sampling may be used at any stage—in the above example the frames may be based purely on geographic distribution, or they may be based on the distribution of population, or some other meaningful category of elements.

Systematic Sampling

The technique referred to as systematic sampling, or the *fixed-intervals* method, is convenient to use when there is available a roster of elements from which selection is to be made and the elements do not lend themselves to randomizing, or when the randomization process would be inappropriate with respect to manpower and finances. The researcher determines the sample interval (every nth element) which will yield a predetermined percentage of the entire roster as a sample. Out of the first n elements, one (k) is drawn at random. Then, counting from k every nth element is drawn thereafter for the sample. If it were appropriate for a given study, for example, to use an alphabetical directory of about 500 names as the population, a 10 percent sample could be collected by drawing at random one name from the first 50, and then every 50th name thereafter.

Insofar as the roster of elements may be considered random with respect to the field of inquiry, a systematic sample may also be considered random. However, rosters are not usually in random order, and one must be careful to use this technique only when the basis on which the roster was compiled and the order of the units on the sample would have no effect on the variables under consideration.

Judgment Sampling

Judgment sampling is based on judgments made by those collecting the data. Two sorts of judgment sampling are *purposive sampling* and *quota sampling.*

In *purposive sampling* the "controls" are usually identified as representative areas (city, county, state, district), representative characteristics of individuals (age, sex, marital status, socioeconomic status, race), or types of groups (school administrators, school counselors, elementary teachers, secondary teachers). These controls may be further subdivided by specified categories within classes; such as the breakdown by amount of training, years of experience, or attitudes toward teaching. In these respects, the controls are somewhat similar to the strata in the stratified sampling method. Usually the researcher develops a master design in which the number of cases to be selected in the sample is proportionate to the total number of such cases with the "control" characteristics in the population; although sometimes these proportions are based on scant and partial information about the population.

In *quota sampling,* an interviewer is given a quota of cases he is to select bearing each of several predetermined characteristics similar to those involved in purposive sampling, but he is given considerable freedom in choosing the individual cases. This may lead to large biases, since those who are often available at the times and places the interviewer meets them are not representative of the population. Oftentimes interviewers have used the line of least resistance in meeting their quotas and have collected information only from those who were easily available for interviews and have ignored people in areas more difficult to reach. Poor supervision of interviewing teams and inadequately defined controls have led to results that have caused serious criticisms of the method.

Some writers believe that the method is particularly weak because, since there is no way of calculating the limits of possible error or the required number for the sample, it does not fit the requirements of probability theory. The same criticism

may be made, however, of any method of sampling in which definite criteria for the selection of the sample are not followed carefully and in which data are collected haphazardly. Since the presidential election in the United States in 1948, when some of the pollsters made grave errors in prediction of results, there has been considerable improvement in the techniques of judgment sampling being used. More attention has been given to developing a good sampling design, better training and supervision have been given to poll interviewers, and a much higher degree of accuracy has been obtained.

Multiphase Sampling

In some research projects it may be desirable to collect certain items of information from a large sample and other items from segments of the original sample. This is often done when a researcher wishes to make an extensive survey of a population with respect to types of data easily and accurately obtainable by a mailed questionnaire method, and then wishes to follow up with interviews of subsamples of the population to investigate some of the more complex aspects of the problem or to corroborate or refine the information already obtained. This procedure has also been referred to as *double sampling,* although it may be extended to more than two samples. The information in the second, or subsampling, phase is usually collected at a later time than the original sample, and the subsamples are usually selected on the basis of information obtained in the first phase.

The essential difference between multiphase and multistage sampling should be noted. In multiphase sampling, the different phases involve sample-units of the same type—all individuals, for example; while in multistage sampling, the sample units are of different types at the different stages—such as, geographical areas at one stage, political divisions at a second stage, and individuals at a third stage. In more complex sampling designs, multiphase and multistage sampling may be combined; that is, multistage sampling may be used to obtain a sampling of individuals according to certain political divisions and geographic

areas, and multiphase sampling may then be applied to the individuals identified by the multistage sampling process.

Replicated Sampling

When a researcher wishes to make detailed comparisons between two or more sets of observations, or wishes to secure information on nonsampling errors such as differences arising from different interviewer biases or techniques, he may use the process of *replicated sampling*. In this process, two or more independent sets of sample units are selected from the same population, and the data from each set are collected by a different investigator. In this manner, an independent estimate of the whole domain is obtained based on the data collected by each investigator. An analysis is then made to determine whether there are significant differences between the sets, or between the investigators, or whether the differences between sets were significant in spite of differences between the investigators. These analyses are discussed more fully in connection with the design of experiments presented in Chapter VIII.

SELECTED REFERENCES

Alexander, Carter. "The Place of Instruction in Library Methods in Educational Research." *Journal of Educational Research* (September, 1935), *29:*104–110. (Emphasizes the need to know library techniques and presents four essential ones.)

Brower, George. "Hunches and Hypotheses," *Michigan Education Journal, 39:*312–313, December, 1961.

Cochran, William G. *Sampling Techniques*. New York: Wiley, 1953. 330 pp. (Refer to index for specific problems.)

Deming, W. Edwards. *Sample Design in Business Research*. New York: Wiley, 1960, 517 pp. (See especially chaps. 1, 2, and 3, and refer to index for specific problems in sampling.)

Gage, N. L. (Ed.). *Handbook of Research on Teaching*. Chicago: Rand McNally, 1963. (See chap. 3.)

Good, Carter V. *Introduction to Educational Research*. New York: Appleton-Century-Crofts, 1959. (Read pp. 72–114.)

Goode, William J., and Hatt, Paul K. *Methods in Social Research*. New York: McGraw-Hill, 1952. 386 pp. (Read chaps. 7, 8, 14.)

Hansen, Morris H., Hurwitz, William N., and Madow, William G. *Sample Survey Methods and Theory*. Vols. 1 and 2. New York: Wiley, 1953. 638 and 332 pp. (A very extensive set of books on sampling methods, theory, and applications. See index for specific references to problems on sampling.)

Mayhew, Lewis B. "The Critical Incident Technique in Educational Evaluation," *Journal of Educational Research, 49*:591–598, April, 1956.

McCormick, Thomas C., and Francis, Roy G. *Methods of Research in the Behavioral Sciences*. New York: Harper & Row, 1958. 244 pp. (See especially pp. 24–44 and 135–176.)

Mouly, George J. *The Science of Educational Research*. New York: American Book, 1963. (See chaps. 4 and 7.)

Slonim, Morris J. *Sampling in a Nutshell*. New York: Simon and Schuster, 1960. 145 pp. (An elementary presentation of the more well-known technical aspects of sampling—well-illustrated.)

Van Dalen, Deobold, and Meyer, William J. *Understanding Educational Research*. New York: McGraw-Hill, 1962. (See chap. 8.)

CHAPTER IV

Observation Techniques in Data Collection

Functions and Types of Observations

Basic Guides to Good Observations
- Obtain prior knowledge of what to observe
- Examine general and specific objectives
- Devise a method of recording results
- Define and establish categories
- Observe carefully and critically
- Rate specific phenomena independently
- Become well acquainted with the recording instrument

Interpretation vs. *Observation*

Schedules and Procedures for Recording Observed Data
- Anecdotal records or behavior diaries
- Periodic summaries
- Check lists
- Ratings and rating scales
- Photographic records
- Time-sample observations

Training Observers

Advantages and Limitations of Observation Techniques

When collecting data for a scientific research study, the researcher will undoubtedly need to observe phenomena himself or use the observations of others. This chapter is concerned primarily with the direct observations of the researcher or a team

of observers whom he has trained. The next three chapters will deal with observations made by others which are made available to the researcher through interviews, correspondence, or the perusal of documentary sources.

FUNCTIONS AND TYPES OF OBSERVATIONS

The researcher's observations may take many forms and have many different functions depending upon his specific purposes and the frame of his investigation. He may manipulate and observe phenomena whose variables he can control as is done in laboratory experiments in the natural and life sciences, or he may observe phenomena over which he has little or no direct control, as is frequently the case in the field of astronomy. He has very little, or no control, for example, over the size of classes in a school experiment, the time of day of certain classes, the kinds of teachers teaching these classes, the physical setting of instruction, and the like. In most educational research it is necessary to make studies of school children where they are in their normal settings. In these instances the researcher must obtain his data directly through his own observations of the phenomena and must perceive what is significant for his purpose.

In the study of various aspects of social phenomena, the researcher may observe individuals or groups of individuals controlled or manipulated by some device or completely uncontrolled. In other words, he may observe the situations to which people are exposed and their responses to those situations. The researcher may act as a member of the group (participant observer) or as an outsider (nonparticipant observer). If he is a participant observer, he commonly lives with, or shares in the life and activity of, the group under study, disguising himself to be accepted as a member of the group. By establishing a favorable rapport with the group, the investigator is able to make observations he would have been less likely to make had he been an "outsider" or a nonparticipant. He may participate in the group for an extended period of time and may therefore be able to collect a wider range of materials than he would be able

to in a series of interviews or short observations. Nonparticipant observations are difficult to make because the presence of an outsider may influence those observed to react somewhat differently than they would if the outsider were not present. However, in many investigations the researcher has been able to obtain usable information after some member of the group has introduced him and assured the others that his presence is acceptable. More extensive details of observational field-work methods, which also have applicability to researches in education, may be found in a chapter by Whyte in *Research Methods in Social Relations.*[1]

The researcher often uses various mechanical devices which can collect data about a group without his being present or which permit him to observe without others knowing he is present. These devices include the one-way mirror through which an observer may see without being seen by those he is observing; sound recorders and photographic equipment which can provide records of events as they occur; and other mechanical, electrical, or electronic devices such as pedometers, fluoroscopes, recorders that indicate which radio or television channels are tuned in upon, radar, and sonar instruments. Infrared photography has been used successfully in several studies to record the reactions of people when normal visibility was impaired or in situations when observations had to be made in darkness.

BASIC GUIDES TO GOOD OBSERVATIONS

Although there are many different observational techniques, some guiding principles are applicable to all of them. Seven principles are discussed briefly at this point and some are given additional attention in subsequent chapters as they apply to other techniques of data collection.

Obtain Prior Knowledge of What to Observe

The researcher must determine in advance what to observe or what types of phenomena merit recording. If he has prior knowl-

[1] Marie Jahoda, Morton Deutsch, and Stuart W. Cook, *Research Methods in Social Relations.* New York: Holt, Rinehart and Winston, 1951, pp. 493–513.

edge of what he is to look for, he will observe and remember more specific features of a situation than he would without such prior knowledge.

Examine General and Specific Objectives

The formulation of the problem and the specific points needing investigation dictate what should be observed and provide some controls over the observer. Ideas of what to observe and record may be obtained from a thorough survey of previous studies related to the general problem. Many investigators make the fallacious assumption that the process of merely looking at phenomena will reveal what is relevant to them and therefore fail to list specific things to look for as a guide in their observations. On the other hand, an investigator should not look only for those things he has determined to observe, but should also "sense" and record other activities which may have a bearing on his problem. Specifying as clearly as possible what behavior, interaction patterns, or other phenomena are to be noted will enable the investigator to make the kind of objective observations and records that will indicate clearly to other researchers the limitations of his data and allow them to verify his research by repeating such observations. The development of an observation schedule will assist the investigator in this respect.

Devise a Method of Recording Results

To conserve time and to standardize the procedures for multiple or independent observations, it is highly important to determine the informational and statistical units to be used in recording observations. Researchers generally agree that observations should be recorded as soon as possible; however, they frequently fail to do this because it is a time-consuming task. To help speed up the recording process, an observation schedule or check list can be designed which will keep the amount of writing to a minimum. If it has been carefully devised, it will include a relatively complete listing of items to be observed as well as space to include other items not anticipated at the time the schedule was developed.

Define and Establish Categories

In addition to making numerical counts of the kinds of behavior to be observed, it is necessary and desirable to make at the same time some qualitative observations and interpretations. Thus, by the use of a two-way table, listing the kinds of behavior on one axis and qualitative ratings or interpretations of them on the other axis, the researcher can readily record much data that would otherwise be difficult to compile. Each category or level of data being collected should be concisely and carefully described by indicating the phenomena the investigator expects to find in each. This description becomes especially important when a team of observers is collecting the information.

Observe Carefully and Critically

By thinking through the problem and determining what to observe, by devising a simple method for recording his observations, and by deliberately concentrating upon the specific aspects which interest him, the investigator will obtain more reliable and authentic data than he will by observing in a haphazard and hurried manner. He must know his subject matter, develop perceptions based on the proposed classification of his data, and develop skill at the process of observing.

Rate Specific Phenomena Independently

Whenever qualitative values are to be observed, each should be rated and recorded independently of others by use of a well-defined rating scale. Frequently investigators are prone to let ratings of one phenomenon influence their ratings on others. For example, in an investigation which involves recording the physical appearance and neatness of attire of teachers and their proficiency in classroom instruction an observer may rate a subject very highly with respect to his appearance, and, because of this, give him a spuriously high rating in other respects. By rating each item to be observed on the defined scale for each, this "halo effect" is more likely to be reduced.

Become Well Acquainted with the Recording Instrument

Familiarity with the observation schedule and any devices used in recording results of observations will enable the researcher to record his observations accurately in a minimum of time and to avoid delays in his work. If he is to use instruments or mechanical aids in making his observations, he should practice with them before collecting data. When a team of observers is to make use of an instrument, preliminary training and periodic conferences and discussions among those making the observations will help to standardize their procedures and increase the reliability and authenticity of the data they collect.

INTERPRETATION VS. OBSERVATION

In recording observations scientifically the researcher should record only what actually happens and avoid introducing any interpretation into the factual data reports *except* as they are necessary *in addition to* the recorded facts. Interpretations and evaluations of the data observed should be withheld until the evidence is all in, although "on-the-spot" interpretations may be noted at the time to assist the researcher when he starts to analyze his data. The difference between an observation that is scientifically recorded and one that is loaded with opinion or interpretation may be seen in the following examples, related by Thomas,[2] of an incident reported by a third-grade teacher.

Observation 1. "Tony did a naughty thing today. He said bad words, swore at some other children on the playground. Then he thought it was funny, and he was quite surly when I reprimanded him about it."

Observation 2. "Tony and Fred and Jane threw sand at each other near the swings after lunch. When Jane and Fred stuck their tongues out, Tony shouted, 'Damn you damn sneaks.' Fred and Jane ran across the yard calling, 'Oh, that's naughty.' Tony shouted after them, 'Ha, ha, run away.' I called Tony to my room. He

[2] R. Murray Thomas, *Judging Student Progress.* New York: Longmans (David McKay & Co.), 1954, pp. 187–188.

looked out the window as he stood at my desk. I said, 'You know, we don't want children using such words as that.' Tony said, 'They threw sand in my hair. Served them right to be called damn.' "

The second observation is a recording of what happened. If the teacher at the end of the term, or perhaps the principal who sees Tony's folder, wants to conclude, "Tony did a naughty thing . . . said bad words . . . thought it was funny . . . was surly," then that is his privilege. But any such personal interpretation should be based upon well-recorded anecdotes of what actually happened. . . . Someone else might well interpret the incident differently, and the behavior should be described accurately to allow such interpretations. . . . Observation 1, above, is colored with opinion and actually does not tell much. In its present form it is a better device for evaluating the teacher's attitude toward Tony than it is a record of an incident in the boy's life.

SCHEDULES AND PROCEDURES FOR RECORDING OBSERVED DATA

Observations may be recorded in several forms depending on their proposed uses. These include anecdotal records or behavior diaries, periodic summaries, check lists, ratings and rating scales, photographic records, recording tapes, and the like. Forms of recording observations are often referred to, in general, as observation schedules. Following are brief descriptions of the different schedules which can be used for recording data in educational research. Modifications of the techniques described may be made to fit specific studies.

Anecdotal Records or Behavior Diaries

Records of either typical or exceptional behavior observed by the researcher or by other individuals are often used in research studies dealing with social adjustment, the effect of environmental changes upon individuals, cause and effect relationships, and the progress of people toward their goals. By keeping such descriptive records over a period of time, the researcher is able to evaluate the significance of behavioral incidents more readily than if he used other recording methods.

However, this method is time consuming, as it must be continued over a period of weeks or months to be valuable. Its effectiveness depends upon the skill of the observer in determining what incidents to record and his ability to describe them by concise and objective word pictures.

Periodic Summaries

In some instances the researcher makes periodic summaries of behavior instead of recording his observations after, or during, each group session. He makes casual observations of the way people act over a period of time and then writes up his general impressions of their behavior. This method is not as likely to be as reliable as the former since many of the incidents which the researcher thought he would remember easily often tend to slip away as time passes. In addition, he is apt to remember more of the nontypical incidents than the typical ones.

A slight variation of this technique has been referred to as the "Critical Incidents Technique." A *critical incident* is an observed behavior or aspect of behavior that is *judged* to make the difference between success and failure in some activity. Commonplace behavior is disregarded in making judgments, and only those acts that are considered to be critical are noted. For example, a school administrator might be asked to recall a teacher's activities over a period of time and to pick out the specific incidents which may have contributed to his success or failure in his job, or teachers might be asked to recall their own college days and remember incidents of effective or ineffective teaching or context that contributed to their success (or failure) at the present time. One of the difficulties with this technique is that it lacks reliability, as it depends upon the identification of behavior that has rarely occurred and may not reoccur—a very limited sampling of observations tending to lack reliability. Also it is difficult to classify and interpret such a small number of incidents, oftentimes quite diverse in nature, and the respondent may identify these incidents in a somewhat questionable manner or may give them higher importance as causative factors than they actually had at the time of occurrence.

Check Lists

A check list is often used in making observations to insure that the observer looks for every bit of evidence that he has previously determined as essential. It should be prepared in advance of the observations, but should include blank space for recording phenomena that were not anticipated in the formulation of the problem. Table 1 is an example of a check list for observing the work habits and skills of pupils. In using this check list the observer would merely check the occurrence of the indicated habits or skills to discover the frequency with which they are demonstrated by the various children under observation.

Ratings and Rating Scales

In addition to the quantitative observations, it is often desirable to make qualitative judgments and to record them at the time of the observation. Rating scales are a condensed method of recording quantitative and qualitative observations. The check list shown in Table 1 could provide considerably more information if used in connection with a rating scale which evaluated the *degree* to which the various work habits and skills were demonstrated. For example, instead of simply using a checkmark to indicate the frequency of occurrence of various habits and skills, the observer could indicate his value judgments by using the rating scale shown in Table 2 and by writing the numerical scale values in the record form shown in Table 1. This example gives the rater only three choices, but six or seven is perfectly feasible and would provide the opportunity for more discriminating evaluation.

Photographic Records

Researchers often use photographs and motion pictures to record observations as this technique permits them to preserve details for more intensive study at a later time. Considerable use is being made at present, for example, of photographs and

TABLE 1. Check List of Work Habits and Skills

Names of Pupils	Comes to Class Ready to Work I	Follows Instructions II	Works Dining Class Periods III	Carries Through Given Tasks IV	Does His Written Assignments V	Prepares for Class Discussions VI	Other			
Allen										
Barry										
Blythe										
Edward										
Fredrika										
(Others)										

motion pictures by those interested in studying traffic accidents and traffic patterns in order to develop safety precautions and by insurance companies in the analysis of fire and other disaster claims. Also large quantities of business records are now put on microfilm so they will be available later for various purposes. By using stills or slow-motion movies the investigator is able to analyze activities or situations that could not be studied at the time they occurred or at the normal speed of their occurrence. They also can be used in training observers to improve the reliability and validity of their observations.

TABLE 2. Rating Scale for Evaluation of Work Habits and Skills (Scale Values Range from 1 to 3 as Indicated)

I. Comes to class ready to work:
   1. Brings work materials; takes his place quickly; gives prompt attention.
   2. Usually brings work materials; takes his place with occasional reminding; often needs to be called to attention.
   3. Rarely brings work materials; takes his place with considerable urging; seldom gives attention.

II. Follows instructions:
   1. Attends to written and oral instructions; follows directions accurately.
   2. Sometimes lets his attention wander; usually follows directions.
   3. Pays little attention to instructions; follows directions reluctantly or not at all.

III. Works during class periods:
   1. Loses no time in getting to work; concentrates in face of minor distractions; works, independently and steadily.
   2. Often loses time in getting to work; reacts to minor distractions; frequently depends on others; works unevenly.
   3. Loses much time; is easily distracted and often disturbs others; does very little work even with close supervision.

IV. Carries through given tasks:
   1. Volunteers for and accepts readily special assignments; prepares carefully and punctually.
   2. Seldom volunteers but accepts special assignments when made; preparation and presentation on time but average in quality.
   3. Never volunteers, and accepts assignments under protest; often fails to have them ready on time or to prepare them adequately.

V. Does his written assignments:
   1. Always completes written assignments on time; has them well organized and in good form.
   2. Generally completes written assignments on time; organization and form acceptable.
   3. Seldom completes written assignments on time; poor organization and form.

VI. Prepares for class discussions:
   1. Uses a variety of sources; takes notes accurately and systematically; speaks to the point, and backs up his opinions with facts.
   2. Usually relies on one source; occasionally has sketchy and disorganized information; and expresses opinions which he cannot substantiate.
   3. Usually comes to class unprepared and attempts to bluff if he expresses an opinion.

Time-Sample Observations

Time sampling consists of recording activities for a definite period at a particular time of day. For example, a fifth-grade teacher might record everything a pupil does for a five-minute period during a free-reading or study session. By this means, evidence can be obtained with respect to a pupil's work or play habits, attention span, and pattern of movements. A different pupil might be chosen another day and watched for the same time. Or, a teacher who desires evidence about the varying patterns of behavior of an individual or of a class at different periods of the day would find time sampling useful. As a tool for research in child psychology, Olson states that this technique:

> . . . is characterized by careful selection and definition of the behavior to be observed, by standardization of the observer's methods, by careful limitation of the length of the period of observation, and by the multiplication of observations to be certain that the behavior of the child has been sampled in an adequate fashion for such variables as situation, time of day, and other factors that may influence behavior. . . . Results that give a reliable basis for experimentation and prediction have been procured. . . .[3]

TRAINING OBSERVERS

The observer himself is always a variable that must be taken into account—an observer is a measuring instrument. As an "instrument" for the collection of data, he must be trained to see accurately what he is supposed to see. Even the person who designs the project may be using poor observation techniques. The following suggestions may be used by the researcher both to improve his own techniques of observation and to train observers to assist him in the collection of his data.

The training of observers, like the development of a research proposal, logically starts with a description of the purposes of the study and of the specific questions under investigation. This enables the observers to carry out their tasks with greater accuracy and to understand why the observation schedule was de-

[3] Willard C. Olson, *Child Development*. Boston: Heath, 1949, p. 7.

veloped in its particular form. The observation schedule should be explained item by item and all questions raised should be carefully answered. Observers should be trained to identify various cues that may be used in some items, to make notations when necessary, and to record their own inferences and interpretations.

Try-outs of the observation schedule should be made by each observer under the direction of the researcher. Then several practice runs should be made using a pilot group similar to the groups to be observed for the research study. The researcher's primary objective is to be assured that uniform practices are developed and carried out in all aspects of the observation process. Thus the practice runs with different observers and the same pilot group may indicate whether the observers are reliable and are performing comparably before the actual study is begun. The observers should also be given an opportunity to make suggestions which may be used in the improvement of the observation schedule or the techniques of observation. Oftentimes a researcher can profit by the assistance of his team of observers if he can use them in the development of the observation schedule.

ADVANTAGES AND LIMITATIONS OF OBSERVATION TECHNIQUES

Whether or not the researcher will be able to use the observational method in a scientific fashion depends on his prejudices and biases, his powers of perception, his outlook, his ability to recognize causal relationships or sequences, and his accuracy in recording the results of his observations. If he can use this method of data collection, he will discover it has certain *advantages:*

1. It is the most direct means of studying a wide variety of phenomena. There are many aspects of human behavior which can be studied satisfactorily in no other way.
2. It demands less of the subjects under observation than other methods.

3. It permits the collection of data in typical behavior situations.
4. It permits the recording of behavior simultaneously with its spontaneous occurrence.
5. It does not depend largely on retrospection or reflection.
6. It allows for the emergence of data that the subjects might not have thought of in interviews or in responding to correspondence. Incidences of everyday occurrence are often neglected in using other methods of data collection.

However, the observation method also has several *limitations:*

1. People, knowing they are being observed, may deliberately try to create favorable or unfavorable impressions on the observer.
2. The spontaneous occurrence of an event often cannot be predicted so that the observer can be present. Many times the period of waiting for spontaneous occurrences is time-consuming and emotionally exhausting.
3. At times unforeseeable factors, such as weather conditions, alternate attractions, and the like, interfere with the observational task.
4. It is limited by duration of events. Some events may spread over a period of years and some may occur simultaneously in different geographic locations making it impossible for the researcher to collect all the necessary evidence.
5. Some occurrences may be reported by subjects through interviews or correspondence which would rarely, if ever, be accessible to direct observation, such as the various private and personal events of people's lives.

SELECTED REFERENCES

Bales, Robert F. *Interaction Process Analysis.* Cambridge, Mass.: Addison-Wesley, 1950. 203 pp. (See index for pertinent readings.)

Festinger, Leon, and Katz, Daniel (Eds.) *Research Methods in the Behavioral Sciences.* New York: Holt, Rinehart and Winston, 1953. 660 pp. (Read pp. 243–296 and 381–416.)

Good, Carter V. *Introduction to Educational Research.* New York: Appleton-Century-Crofts, 1959. (Read pp. 221–231.)

Goode, William J., and Hatt, Paul K. *Methods in Social Research.* New York: McGraw-Hill, 1952. 386 pp. (See chap. 10.)

Mark, Mary Louise. *Statistics in the Making.* Columbus, Ohio:

Bureau of Business Research, Ohio State University, 1958. 436 pp. (Read pp. 72–79.)

Selltiz, Claire; Jahoda, Marie; Deutsch, Morton; and Cook, Stuart W. *Research Methods in Social Relations.* Rev. ed. New York: Holt, Rinehart and Winston, 1959. 622 pp. (Read especially Chaps. 6 and 10.)

Travers, Robert M. W. *Introduction to Educational Research.* New York: Macmillan, 1958. 466 pp. (Read pp. 161–230.)

CHAPTER V

Techniques of Interviewing in Data Collection

Functions and Types of Interviews

Basic Guides to Good Interviewing

- Determine who is to be interviewed
- Make preliminary arrangements for the interview
- Determine the plan of the interview and the questions to be asked
- Conduct a preliminary try-out of the interview plan
- Become familiar with various interviewing processes and techniques
- Check the accuracy and reliability of the information obtained
- Make a written report of the interview as soon as possible

Advantages and Limitations of the Interview Method

In much educational research, some information must be collected in the field through personal interview. Certain facts and opinions can, of course, be obtained by mail or telephone, but some information can only be secured in a face-to-face interview. It is often necessary to see one another, to hear one another's voices, to listen to one another's words, and to use all that is psychologically inherent in physical proximity. An interview is usually a nonreciprocal relationship between the individuals concerned; that is, one party desires to get information—one party *interviews* the other—for a particular purpose.

FUNCTIONS AND TYPES OF INTERVIEWS

The interview must be embarked upon with a sense of its functional unity. It must have a definite objective and not be just an occasion for unorganized, disjointed observations and judgments without beginning or end. The investigator conducting the interview has three main tasks: informing the respondent about the nature of the project and explaining why his cooperation is desired, motivating the respondent's interest so that he will cooperate, and, most important, obtaining information. Picking up documents, statistical reports, and the like, and locating new sources of unrecorded data are all part of the latter task. An interview affords opportunities to observe individuals and groups in action and to learn facts, opinions, and beliefs which may vary with particular persons in particular circumstances. It may serve to corroborate data already obtained from various independent sources or to disclose apparent contradictions or other discrepancies among sources. It can aid in the critical evaluation of other sources of data and act as a check on the reliability of data obtained by other techniques. However, it should *not* be used to compile data of uncertain value or to get information more readily available elsewhere.

Interviews are sometimes described or classified according to their purpose. The most common types are survey, diagnostic, therapeutic, and counseling. *Survey* interviews are used to obtain information from persons considered to be authorities in their fields or representative of groups about which information is desired. This type of interview is commonly used in making political polls, in surveying attitudes about educational programs or school personnel, in determining views of teachers toward administrative policies, in gathering opinions from firms and the public, and in collecting information on other problems which may affect policy and decision-making matters in educational institutions. The objective of the *diagnostic* interview is to understand a problem, the causes that gave rise to it, its present status, and its seriousness. The major purpose of the

therapeutic interview is to help an interviewee to understand himself better and to plan appropriate therapy. This type of interview seeks to eliminate, or ameliorate, casual factors and to promote the improvement of the interviewee's emotional life. The aim of the *counseling* interview is to enable the interviewee, working together with the counselor, to gain greater insight into personal or vocational problems and to make sound plans to solve these problems. It is obvious that of these four types of interviews the survey interview has the greatest application for research problems in the field of education, although some problems may require the use of the other types.

BASIC GUIDES TO GOOD INTERVIEWING

When collecting information through a personal interview, it is necessary to have a plan which takes into account not only the data required but also the personalities of the researcher and the individuals to be interviewed. The following guides should be followed in planning and carrying out interviews.

Determine Who Is to Be Interviewed

The interviewer, since he seeks well-authenticated facts, or at least informed judgments, must select the individuals he will interview with the greatest care. He must determine which people have the necessary information, whether or not they have authority to divulge it, and, if they do, if they will be willing to do so. One of the worst mistakes a researcher can make is to misclassify a respondent—to interview a respondent who is not able to furnish the requisite information. The researcher should attempt to find out as much as possible about the interviewee's responsibilities, present and past business connections, possible biases, experiences in the field, and so forth. It is also important for him to determine at the outset how many people he should see in order to obtain sufficient information to serve as the basis for reliable generalizations and whether or not these people are representative of all the factions or subdivisions of the group being studied.

Make Preliminary Arrangements for the Interview

A definite time and place should be set for the interview. The researcher should set a time that is satisfactory to the interviewee and meet the appointment promptly.

In most educational research the interview will probably occur at the respondent's home or place of business, so the interviewer can do little to control the physical surroundings of the interview. As preliminary preparation for such interviews, the interviewer should learn something about the regular pattern of the respondent's activities and avoid rush periods of the day, week, or month. If it is possible to influence the circumstances of the interview, the interviewer should suggest privacy to encourage confidential statements and uninterrupted quiet so the informant's interest, once established, can be maintained until all information desired has been obtained.

However, if it is impossible to influence or control the situation—as when an interviewer must collect information on downtown street corners where he may be distracted by adverse weather conditions or other things, or in homes where he is interrupted by telephone calls, children, and other domestic activities—the investigator must realize that the data obtained are not likely to be as satisfactory for research purposes as those obtained under controlled situations. In informal and uncontrolled situations, the interviewer should still try to avoid rush periods in the day. He should select times when children are in school or evening periods when there are likely to be fewer adverse conditions.

Occasionally it is desirable to facilitate contact through mutual friends or acquaintances, letters of recommendation, or personal letters. Some interviewers have found that the use of credentials or letters of introduction is helpful in making contact with the interviewee.

Determine the Plan of the Interview and the Questions to Be Asked

In preparing for the interview, the researcher should decide just what is to be accomplished, what facts are to be brought

out, what information is to be solicited, what attitudes are to be established, and what actions are to be motivated. In order to do this he must know his field and have his problem well formulated.

The interview may be either an exploratory one or one to collect information. The former is relatively unstructured and serves as a gateway to many sources of information: it may enable the researcher to obtain expert advice on technical aspects of the field, to corroborate data from other sources, and to evaluate information. The second type should be highly structured and make use of a previously prepared schedule or list of questions. Even the data collection interview, however, should be flexible enough so that the investigator can adapt to various types of informants and record information he had not anticipated.

Conduct a Preliminary Try-Out of the Interview

Before carrying out an actual interview, the researcher should conduct interviews with friends, classmates, or other persons *not* in the interviewing sample but like those from whom the research information is desired. This may help the researcher improve his technique of asking questions and should provide some indication of the kind of responses he is likely to obtain. In general, the researcher should be careful to make pilot runs of all techniques he anticipates using.

Become Familiar with Various Interviewing Processes and Techniques

One of the interviewer's major objectives is to elicit the interest and cooperation of the respondent. Since he expects frank and, perhaps, personal answers, he must be frank and open with the respondent regarding the purpose of the interview. Successful interviewing is a true art for which no hard and fast rules can be given. Refusals can wreck a carefully selected sample, and inadequately trained interviewers are much more likely to encounter refusals than are experienced ones.

The researcher should become proficient at: (1) creating a friendly atmosphere, (2) asking questions, and (3) obtaining

responses. How he should do this is discussed in detail in a number of books,[1] but some of the basic rules of interviewing follow.

In establishing friendly relations, attempt to develop a cordial setting for the interview. Engage in pleasant conversation, through which questions can arise naturally. Try to establish a relationship of mutual confidence by stating the purpose of the interview clearly and simply with sincerity and reasonableness, linking the topic of the inquiry to the interests of the person being interviewed. Establish pleasant associations or circumstances. Help the interviewee to feel at ease. Be at ease yourself and make it apparent. Allow him time to get acquainted with his surroundings and to gain poise. Make him feel that he is an equal who can exchange ideas and opinions with you. Avoid letting the interview take on the character of an inquisition or "third-degree," and also avoid making it an oral "questionnaire."

In getting to the essential points of the interview, begin with pleasant topics and avoid unpleasant associations and emotional attitudes. Ask questions at first that will not provoke any form of negativism and which the interviewee will not refuse to answer. Start with emotionally neutral topics; gradually lead into more personal and emotionally colored questions as the interview progresses. Be straightforward, rather than shrewd or clever. The interviewee is likely to detect any cleverness or deviousness on the part of the interviewer and to answer in like manner. Approach the problem as soon as rapport is assured, but do not ask pertinent questions until you think the interviewee is ready to give the information accurately and willingly.

To facilitate responses, practice taking the interviewee's point of view—think in his place. Ask only one question at a time, and avoid implying the answer to your own question. Focus the interviewee's attention on the question, and make sure that he understands what is wanted. Listen, give him a chance to talk, let him tell his story, and then help him to

[1] Note references of Bingham and Moore, Fenlason, and Garrett at end of this chapter.

supplement it. Observe his behavior, facial expressions, tones of voice, and other indirect responses. Allow time enough for him to answer one question fully, but prevent him from rambling by keeping control of the interview without being domineering. Keep on the subject, but listen carefully for casual remarks that might be revealing. If unusual facts are revealed, do not betray surprise, shock, or emotional tension at the disclosures. Face the facts professionally, and avoid excessive sentimentality, sympathy, or antipathy.

Avoid taking the role of "teacher"; avoid a patronizing attitude; do not embarrass the interviewee unnecessarily; avoid putting him on the defensive or arousing feelings of antagonism; and keep him from making the interview merely a social visit. In phrasing questions, be careful not to offer alternative questions that might imply what you think is the correct answer—do not "put words in his mouth." If the interview runs longer than the appointment schedule of the person allows, arrange for another interview at a later time, rather than extending the present session. If the interviewee knows other people are waiting to see him, he may want to conclude the appointment and may answer superficially.

A final word should be said about the interviewer's personal appearance. Ordinarily, he should be neat, well-groomed, and conservatively dressed. However, in some circumstances (*e.g.*, in participant-observation types of investigations), he should endeavor to appear and conduct himself as though he himself were in the same peer group as the respondent.

Check the Accuracy and Reliability of the Information Obtained

As responses are made to questions raised in the process of the interview, the interviewer should consider their accuracy. There are several sources of errors. They may occur as a result of defects in sense organs, especially those of sight and hearing. The interviewee may make errors when estimating time and distance; he may underestimate long distances or long periods of time and overestimate short distances and short periods of

time. When people respond to questions about incidents or situations requiring long spans of memory, the responses are likely to involve errors. Likewise, some persons tend to overrate their enjoyment of pleasant experiences and underrate their dislike of unpleasant occurrences.

Since some people tend to exaggerate, to make loose or implausible statements, or to practice deliberate deceit, it is necessary to be on the lookout so as to minimize these tendencies. In checking on a statement of objective fact, the interviewer may be able to suggest it will be further investigated and thereby make the respondent more careful in subsequent statements. Watch for indirect ways to verify indications of preference or attitude; observe facial expressions, listen to tones of voice, and so on. Help the interviewee to realize it is his responsibility to be accurate. Be sure he has understood the questions and that you have understood his answers. Give him an opportunity to qualify his responses. Check his answer at once, whenever possible, by repeating it in your own words and asking if that is what he meant.

Practice separating facts from inferences. Check statements of fact which tend to reflect some measure of emotion to see if they were influenced by your own prejudices or those of the interviewee. If percentages or fractions are presented by the respondee, translate them into numbers and ask him if that is "about right." For example, a respondee may state that he works at a given task about 20 percent of his working time. By asking if that means about an hour and a half a day, the interviewer can check the basis of the percentage figure given and determine the accuracy of the response. At the end of the interview, it is desirable in some instances to secure a confirmatory written summary. This should be prepared by the interviewer and submitted to the interviewee for confirmation.

Make a Written Record of the Interview as Soon as Possible

To insure the greatest accuracy, it is desirable to record, at the very earliest opportunity, all data and information obtained

during the interview. Notes may be taken during the interview, or they may be recorded immediately following it. The practice of note-taking is subject to individual habits and predilections. If notes are taken during the course of an interview, it is essential that the interviewer know how to combine the process with the conversational aspect of the interview, as long pauses in the conversation for the purpose of writing notes are usually not to be desired. On the other hand, writing notes while the interviewee is talking may cause him to cease talking until the interviewer is through writing. Many people object to having a record made of what they are saying, and the interviewer must be able to decide in each case if his note-taking is likely to prejudice the interview and to modify his practice if it appears desirable to do so.

In general, it is foolish to attempt to conceal the fact that notes are being taken. Some interviewers think that they can "doodle" and fool the interviewee, but this is usually an erroneous belief. As a rule, the interviewer should indicate to the respondent that his answers are important, that it is desirable to record them as completely and accurately as possible, and should request permission to take notes during the interview. Unless the interviewee is strongly opposed to note-taking, the process should be practiced. It allows the interviewer, at times, to take his eyes off the subject without appearing to let his attention wander; this is advantageous as the interviewee may become disconcerted if he feels constantly regarded.

It is desirable, in planning for the interview, to develop an interview schedule or questionnaire that is precoded in such a way that the interviewer can check the code that comes closest to the interviewee's response. This record blank should provide space, in addition to the predetermined code, to write in brief remarks that were not expected at the time it was developed. Immediately following the interview, the record should be inspected to make sure it has been filled in accurately and completely. If there are items for which free responses are requested, the interviewer should make complete verbatim reports, if at all possible. These verbatim reports should not

be edited or changed, even if the respondent's remarks include poor grammar, profanity, or slang expressions, for it is from these remarks that attitudes, emotions, and veiled meanings may be inferred.

A complete transcript, whether obtained by wire and tape recordings, stenographic notes, or other means, is the most reliable and valid record. Unfortunately, the use of mechanical devices may give an air of stiffness and formality to the interview which tends to inhibit frank discussion. In addition, these devices are not able to record indirect responses such as facial expressions, shrugging and other body movements, intonations of voice, and other nuances which may play an important role in the interview process. These need to be noted directly by the interviewer.

It will be well to consider at this point some of the kinds of errors that often occur in reporting interviews. If the interviewer fails to recognize or tends to overlook or to minimize significant events, he commits an error of *recognition*. If he, as is almost always the case, omits some fact, expression, or experience, he commits an error of *omission*—an error less likely to occur in the coded or mechanically recorded reports than in those based on note-taking. If the investigator elaborates or exaggerates the respondent's remarks, he commits an error of *addition*—an error less common than that of omission. If he does not recall exactly what was said and substitutes words having different connotations than the ones used by the respondent, he commits an error of *substitution*. And finally, if he does not recall the proper sequence of events or the proper relation of facts to each other, he commits an error of *transposition*. The fact that these errors can easily occur explains why we have stressed the importance of accurate reporting.

ADVANTAGES AND LIMITATIONS OF THE INTERVIEW METHOD

The extent to which a researcher is able to use the interview method in a scientific fashion depends on (1) the abilities of the interviewer to make use of conversation, gauge expressions,

perceive attitudes, encourage formation of judgments, and take advantage of favorable circumstances; (2) his proficiencies in analyzing the essential points of the interview; and (3) his accuracy in reporting the interview. Subject to these considerations, there are certain *advantages* of the method itself:

1. It is the method best suited for the assessment of personal qualities.
2. It is of definite value in diagnosing and treating emotional problems.
3. It is of great use in counseling.
4. It provides information to supplement other methods of collecting data.
5. It may be used, together with observation, to verify information obtained through correspondence methods.

However, it has certain *limitations:*

1. It takes a good deal of time and energy and may be expensive for the researcher.
2. Its success is dependent upon the interviewee's willingness to report and his ability to report accurately.
3. It is influenced by stresses, strains, and other factors affecting either the interviewer, the interviewee, or both.
4. It is influenced by the emotional "set" of the interviewee—by his self-concern or self-pity, by his desire to appear at good advantage, his hesitancy to report facts uncomplimentary to himself, his urge to please or antagonize the interviewer. To some extent any interviewee will color and distort the facts he discloses.

SELECTED REFERENCES

Bader, Carolyn F. *The Interviewer's Guide.* St. Louis, Mo.: Institute of Market Research, 1947. 110 pp. (Several pertinent interviewing techniques and mistakes illustrated by cartoons.)

Bingham, Walter Van Dyke, and Moore, Bruce V. *How to Interview.* 3rd ed. New York: Harper & Row, 1941. 263 pp. (See especially chaps. 1, 2, 3, 11 and 12.)

Fenlason, Anne F. *Essentials in Interviewing.* New York: Harper & Row, 1952. 352 pp. (Refer to index for specific readings.)

Festinger, Leon, and Katz, Daniel (Eds.). *Research Methods in the Behavioral Sciences.* New York: Holt, Rinehart and Winston, 1953. 660 pp. (Read especially pp. 327–379.)

Garrett, Annette Marie. *Interviewing: Its Principles and Methods.* New York: Family Welfare Association of America, 1942. 123 pp. (Refer to index for specific readings.)

Goode, William J., and Hatt, Paul K. *Methods in Social Research.* New York: McGraw-Hill, 1952. 386 pp. (See chap. 13.)

Kahn, Robert L., and Cannell, Charles F. *Dynamics of Interviewing.* New York: Wiley, 1957. 368 pp.

Merton, Robert K.; Fiske, Marjorie; and Kendall, Patricia L. *The Focused Interview.* Glencoe, Ill.: Free Press, 1956. 186 pp. (Excellent reference for techniques of depth interviewing.)

Oldfield, R. C. *The Psychology of the Interview.* London: Methuen, 1941. 159 pp. (Read chaps. 1, 2, 4, 5, and 6.)

Parker, Clyde A., *et al.* "Questions Concerning the Interview as a Research Technique," *Journal of Educational Research, 51:*215–221, November, 1957.

Parten, Mildred. *Surveys, Polls, and Samples: Practical Procedures.* New York: Harper & Row, 1950. 642 pp. (Read especially pp. 82–85 and 331–421.)

Payne, Stanley L. *Art of Asking Questions.* Princeton, N.J.: Princeton University Press, 1951. 249 pp. (Refer to index for specific readings.)

Selltiz, Claire; Jahoda, Marie; Deutsch, Morton; and Cook, Stuart W. *Research Methods in Social Relations.* Rev. ed. New York: Holt, Rinehart and Winston, 1959. 622 pp. (See especially chap. 7 and Appendix C.)

Young, Pauline V. *Scientific Social Surveys and Research.* 2nd ed. Englewood Cliffs, N.J.: Prentice-Hall, 1949. 621 pp. (Read pp. 199–219 and 243–264.)

CHAPTER VI

Questionnaire and Correspondence Techniques for Data Collection

Misuses of the Questionnaire Technique

Questionnaire Design and Format
- Types of questionnaire items
- Structured vs. unstructured items
- Basic guides to item construction
- Directions for responding
- Questions to establish favorable respondent rapport
- Length of the questionnaire
- Design to facilitate tallying procedures
- Questions to test validity of responses

Techniques to Facilitate Usable Returns
- The covering letter
- Use of return envelopes and postage
- Use of a participant reply card
- Follow-up procedures
- Timeliness
- Professional appearance of the questionnaire
- Group questionnaire technique

Pretesting the Questionnaire

Preliminary Questionnaire Analysis

The legitimate use of mailed questionnaires and personal correspondence to collect data is limited to opinions, preferences, and *facts known* to the individuals answering them. As long as the respondents' opinions are with respect to their

preferences, they may be quite valid, but opinions about facts are utterly worthless unless these opinions are, in themselves, the focus of the research. However, facts may be elicited in many situations and for many purposes by the proper construction and use of questionnaires. Questionnaires are used primarily in making status studies of current practices and in conducting opinion polls and surveying attitudes.

The term *Questionnaire* as used in this chapter is restricted to a data collection instrument or schedule to be filled out by an informant rather than by the researcher. When using either the observational or interviewing techniques, the researcher makes out the schedule and collects and records the data on it. Only incidentally does an informant ever see it. In face-to-face situations the researcher can explain the purposes of his investigation, he can arouse the interest and the cooperation of his informants, and he can be assured that they understand any questions asked and are giving reliable answers to them. In using a questionnaire, however, the researcher can accomplish the same thing only by skillful construction of the questionnaire and covering letter. The researcher must assume that his informant is a competent source of data and will provide it willingly. He must also assume that his informant has the ability to understand the questions asked, as intended by the researcher, and that he has answered them in the form intended and with integrity. If the questionnaire technique is to provide valid data for the investigation, the researcher must construct his questionnaire so as to elicit reliable and authentic information.

MISUSES OF THE QUESTIONNAIRE TECHNIQUE

The questionnaire technique, although a popular research tool, has been badly misused. A few of the more common misuses follow:

1. Requests for information which is readily and more accurately available from other sources. Continued use of questionnaires for these purposes annoys the respondent so that he may become unwilling to fill out any questionnaire.

2. Failure to create an incentive to respond.
3. Inclusion of questions a respondent might consider ridiculous or unimportant.
4. Inclusion of questions which, because they suggest or because they otherwise encourage responses which favor the respondent, may result in misleading answers.
5. Inclusion of equivocal or ambiguous questions.
6. The use of questions to be answered by "yes" or "no" or specified multiple answers, when the respondent cannot be expected to answer without a considerable amount of explanation. Answers to be checked make for easy tabulation, but may lead to nonresponse or incorrect response.
7. Promises and commitments made to respondents not subsequently fulfilled.
8. Using a questionnaire whose form and length may tend to discourage busy respondents.

With respect to the first, and probably the most frequent single misuse, the following "open letter to a student" was published by Heberle.

Dear Sir:

In reply to your letter and questionnaire concerning your research project, I would like to give you my candid reaction to the kind of procedure which you propose to use. I assume that your major professor knows about your questionnaire and I want him to read this letter. If your request were singular in its kind, I might not take the trouble, but it seems to represent a pattern of graduate "research" technique which is spreading—a pattern which in my opinion involves not only unnecessary imposition on the time of busy professors, but also a decline of scholarly standards.

First of all, you are asking in the questionnaire certain questions you could easily answer for yourself by consulting *Who's Who in America* or the university catalogues.

More serious is the objection I have to the rest of your questions; you are asking me for my opinion on very complex questions, and you formulate your questions in a way that indicates you expect a dogmatic answer. To do real justice to these questions, which concern the objectives and methods of . . . sociology, I would have to write you an essay, or several papers. It is hard to imagine that

you really expect me to do this for you; if you do not, then why ask these questions? Furthermore, it so happens that I have expressed my ideas on these matters in several publications; I admit that my opinions are in some cases not stated explicitly but by implication. Now there is an old and well-established way of getting information about other scholars' opinions and theories; that is, by reading and critical interpretation. There is no substitute for this. My advice to you is to forget about the questionnaire and to study the literature.

Very sincerely yours,

Rudolph Heberle[1]

Examples of the kind of undesirable questionnaires referred to by Professor Heberle are shown in Figures 4 to 6. The "shot-gun

Dear sirs;

I am a student at State Teachers College.
I am compiling pamphlets and bulletins
related to your activities for a teaching
file in order to receive a degree. Any
information you may have to send will be
of great value to me. Thank you.

FIGURE 4. A "Shotgun" Postal Card Request.

request," shown in Figure 4, illustrates a common practice—the soliciting of information in very general terms. Frequently requests such as these are sent out by postcards, which makes them even more impersonal. It is evident that the individual sending out this request has very little understanding of

[1] Rudolph Heberle, "On the Use of Questionnaires in Research: Open Letter to a Graduate Student," *American Sociological Review*, August, 1951, p. 549. (Reprinted by permission of the author and the *American Sociological Review*.)

the problem about which he is inquiring. He has made little or no detailed analysis of his problem and is simply seeking information by a "shotgun" method. There is no way in which the recipient could determine the researcher's specific objectives in order to send him the kinds of information that would be of most value to him. This is the type of request that irritates recipients and creates an aversion to the questionnaire technique for collecting data.

In Figure 5 the researcher has evidently been "spoon-fed" a problem and does not have much of an understanding of it. It is highly doubtful that he would receive a negative response

Dear Fellow-Teachers:
Mr. A, State Supervisor of Occupational Information and Guidance, has suggested that, in order to furnish his department with necessary information and to satisfy partial requirements for a Master's Degree, I submit the questions on the attached card to each High School Principal in the State. We are very earnestly hoping for an answer from every school.
If you will mark these questions and return the attached card as soon as possible, your helpful cooperation will be sincerely appreciated.

Does your school provide for every student to have an individual Educational and Vocational Plan?__________(Yes or No)

If not, do you think it should provide this opportunity?____________(Yes or No)

____
Signature

____
School

FIGURE 5. Return-Reply Card Request for a "Spoon-Fed" Problem.

to either of the questions he asks. It is difficult for the author of this text to see how the information requested can provide a basis for a master's thesis unless there is an extensive follow-up to collect additional information.

Figure 6 represents a much better designed questionnaire in which a problem is stated and the information required appears to be pertinent. It should be noted that the researcher

As part of my work toward fulfilling the requirements for a Master's Degree in Education at the University of ______, I am undertaking a survey of some curriculum practices that relate to general education in the social studies in teacher training institutions. Your answers to the questions on the self-addressed return card will be most helpful in studying curricular patterns and will be greatly appreciated.

Very truly yours,

1. Name of School. U. of Oregon
2. School grants bachelor's degree or equivalent? Yes-No
3. School emphasizes training for (underline): elementary school, secondary school, both?
4. To what extent does school set requirements in areas of general education for all trainees? Little-Much
5. Are basic general education courses in the social studies required of all trainees? Yes-No
6. If these are not required, are electives available in social studies?
7. Some course labels used (underline): Problems of Democracy, Modern Society, Western Civilization, American History, Survey of Social Science, other.
8. Name of staff member working with social studies (general education) program.

Signed:

FIGURE 6. "Seeking Unnecessary Information" Type of Postal Card Questionnaire.

has identified the respondent by writing the identification information on the first line of the postal card reply card. However, this one is illustrative of a large number of questionnaires that are sent out without justification. The information for all but one of the questions (Number 8) is available in documentary sources—college catalogues and other printed materials usually provided by institutions. These would be more reliable sources for the information than the memories of the individuals responding. If the name of the staff member acting as advisor for undergraduate students (see question Number 8), is not provided in an institution's publications, a postal card requesting his name would be satisfactory. If the institutional publications had been surveyed prior to sending out the questionnaire, only a letter would have been needed to solicit additional information.

The covering letter and questionnaire, illustrated by Figures 7A and 7B, indicate that the researcher attempted to work on

Dear Sir:

I am in need of data concerning the qualifications of science teachers that get teaching certificates from colleges and universities. Information as to the need for applied science in addition to regular academic courses is my main interest. I am doing work in determining whether the present-day science teacher is qualified to teach in this rapid moving atomic age.

Yours truly,

FIGURE 7A. Inadequate Covering Letter.

a problem without an adequate background as revealed by requesting information from colleges that should have been requested from state departments of education. In no state in the United States are teaching certificates issued directly by colleges or universities, but by the State Department of Education. In addition, practically all the information requested is available in either the publications of certification requirements issued by state departments of education, or college and univer-

sity catalogues. Another criticism of this questionnaire is in its manner of reproduction. It was so poorly reproduced by a liquid-process duplicator that it had to be retyped before it was printed in this book. The faculty advisor for this student should

QUESTIONNAIRE

Check and complete the following requirements for teacher certification and courses offered for the advancement of science teaching in your state.

Pre-service education	Elementary	Secondary
1. Minimum scholastic requirements (Cr. hrs)	______	______
2. Requirements in student teaching (Cr. hrs)	______	______
3. Physical examination required	______	______
4. Minimum age	______	______
5. Professional course requirements	______	______
6. Comprehensive teaching area requirement (Cr. hrs)	______	______
*7. Applied science requirement	______	______
*8. Technique in the operation demonstration and servicing scientific aids and apparatus	______	______
*9. Courses offered in the teaching of science	______	______
In-service education		
*10. Science workshop	______	______
*11. Graduate science courses offered to meet the needs of the science teacher instead of the research scientist	______	______

*Data that will be of special interest. Your consideration will be greatly appreciated.

These findings will be forwarded to P. G. Johnson, Specialist for Science Division of elementary and Secondary schools, Federal Security Agency, Office of Education, Washington, 25, D.C.

FIGURE 7B. An Unnecessary Questionnaire Sent to the Wrong Recipient.

never have permitted him to send out this questionnaire and covering letter in its present form.

Figure 8 illustrates a method of obtaining information from a letter alone. This particular letter, however, represents some of the common faults of the unstructured type of item, which are discussed later in the chapter. The researcher asks for a "description of the extended activities" of the teacher training program of a university. The problem is not stated specifically, nor is there any indication of the kinds of information the respondent

Director of Supervised Teaching
School of Education
University of ________
City, State

Dear Sir:

I am enrolled as a graduate student for the summer session at the University of ________. I am writing a graduate paper on the topic, Laboratory Experiences for Students Preparing to be Teachers. I would appreciate a description of the extended activities program at your university. Please send me any available material concerning this subject.

Your promptness and attention to this matter is of utmost importance to me. Thank you.

Sincerely,

FIGURE 8. A "Fishing" Letter.

is to provide. It is apparent that the student in this case is looking for a "handout" of materials with the hopes that he can find something that can be "put together" for a term paper or thesis.

In using the questionnaire technique, it is necessary first of all to decide what facts or opinions are to be solicited. What has been said in this respect in the previous discussion of observation and interview techniques is directly applicable to the questionnaire technique. The researcher must determine the persons to whom the questionnaires are to be sent and must write a covering letter to enlist their cooperation. In addition to the considerations presented in this chapter, nearly every prin-

ciple described in Chapters IV and V for the conduct of observations and interviews should also be observed in administering questionnaires.

QUESTIONNAIRE DESIGN AND FORMAT

The arrangement and appearance of the questionnaire is of prime importance in getting good returns. The format should be attractive, as easy on the eyes as possible, and there should be a minimum of difficulty in passing from one question to another and in filling in the intended responses. A well-designed questionnaire will appear to be more important and professional than one that is poorly reproduced and crowded, and it will tend to elicit a higher percentage of valid returns.

Every questionnaire should include, either in its body or in a covering letter, (1) a descriptive title for the study, (2) a brief statement of the purpose of the study, (3) the name of the sponsoring agency or institution, and (4) the name and address of the person to whom the completed questionnaire is to be returned.

Types of Questionnaire Items

All questionnaire items may be divided into two main classes: (1) those for which the respondent *supplies* the words, numbers, or other symbols which constitute the responses, and (2) those for which the respondent *selects* responses from among those presented with the items. The major forms of items may be classified as free-response, open-end, or short-answer, which represent the *supply type;* and the yes-no, true-false, or multiple-choice, which represent the *selection type*.

Free-Response, Open-End, or Short-Answer Form. This form is characterized by the presence of a blank on which the respondent writes the information called for by the directions. It allows the respondent to write how he feels about a topic and to give the background of his answer. It is also useful for obtaining information that cannot be classified into specific categories until after the data have been received, and for obtaining data that might require too large a number of categories.

This type of item is often difficult to tabulate and summarize because of the variety of different answers that may be given by respondents. It is almost impossible to phrase free-response questions on certain essential topics so that the same responses will be made by all those from whom information is sought. In addition, a respondent may inadvertently overlook many things that he would have reported if he had thought of them or had been reminded of them. Researchers have often been surprised at the variety of correct responses to questions which they thought had only one answer. Respondents may fill blanks of short-answer items with words which are appropriate, but which fail to express their actual intent. The analysis of such replies is less objective, more cumbersome because of the variety of categories presented, and more time-consuming. And the resulting summaries and conclusions are less meaningful.

Yes-No, Right-Wrong, or True-False Item Form. This form consists of statements to be answered categorically as yes or no, right or wrong, true or false. It is essentially a form in which only one of the possible alternates is explicitly stated.

To meet the standards of objectivity, a statement must be so precise in phrasing and so universal in application that it requires no additional qualifications and admits of no possible exceptions. This requirement tends to limit the applicability and validity of items of this type since many responses cannot be made accurately as either "black or white," but must be made in some shade of "gray." Even qualifying the responses as being "usually yes" or "usually no" does not avoid a degree of ambiguity.

Multiple-Choice Form. This form consists of an introductory statement or question and three or more suggested answers. It has fewer of the weaknesses of the preceding forms and is adaptable to a wide variety of questions. In this form the respondent merely checks the suggested answer that seems most accurate or best applies to him. Sometimes it is desirable to provide a blank space for the respondent to write in a response to a question if none of the suggested answers seems to him to apply. The form is also adaptable to "position marking" on a

separate response form similar to the separate answer sheets used at present for many standardized tests and inventories. The multiple-choice form is the easiest to use and the most objective for tabulating and summarizing responses. Various mechanical techniques[2] using "keysort" and "data-processing" equipment have been developed to facilitate multiple-choice questionnaires, and their use should be considered at the time the questionnaires are planned.

Structured vs. Unstructured Items

The use of structured or unstructured approaches to data collection has been a controversial subject among research workers. Some persons contend that the structured form is likely to force the respondents and to "put words in a respondent's mouth." On the other hand, the unstructured form is suseptible to inadvertent omissions on the part of the respondent. The structured form of items for collecting highly objective data that can be obtained from record sources, or data that do not require a respondent to make value judgments, has not received as much adverse criticism from the proponents of the unstructured form. The more subjective the nature of the information sought, the greater is the need for extreme care in the formulation of items, whether they be in a structured or unstructured form. The decision as to the form to use for any research problem must be made by the researcher himself, and it should be made on the basis of his particular investigation. Figures 9 and 10 illustrate similar types of items in unstructured and structured form.

The unstructured item often requires the respondent to do some hard, reflective thinking, and may necessitate a lengthy discussion on his part. In the process he might unintentionally omit some of the factors just because he did not happen to think of them at the time. Since he probably has nothing personal to gain immediately from answering such items, and probably is a stranger to the research worker making the investigation, the chances of his answering the questionnaire could be slight.

[2] For details in using these techniques, refer to Chapter IX.

1. What extracurricular or leisure-time activities do you believe should be provided for the youth in your community?____________________

____

____

2. Which should be sponsored by the school and which by community agencies?____________________

____

____

3. What do you consider the value of each activity?

____

____

____

FIGURE 9. Example of Unstructured Items.

The structured item requires only a checking or writing in of a scaled value judgment, and also enables one to ask several more specific questions about the same list of activities. It should be possible to provide a few extra blanks in each section for writing in additional ones that might occur to the respondent but which the researcher did not think of at the time he prepared the questionnaire.

Basic Guides to Item Construction

The preparation of good questionnaire items is one of the most important tasks of the researcher. When one cannot readily call, in person, on all the individuals from whom information is desired or where there seems to be no need to do so, the reliability and validity of the data obtained must depend upon the adequacy of the questionnaire. The researcher should ask himself the following questions: Why should the respondent answer the questions I have asked? Does he have the information necessary to answer the questionnaire? Have I given him good reasons why the questionnaire should be answered? Have I asked the questions in such a way that there can be no other

In column I, check those activities that are carried on under the direction of your *school.*

In column II, check the items sponsored by *community* agencies.

In column III, check the items you think *should be sponsored by the school* regardless of whether they are being sponsored at present by the school, community agencies, or neither.

In column IV, check the items you think *should be sponsored by community agencies* regardless of whether they are being sponsored by the school, community agencies, or neither.

In column V, rate the value you think the following activities have for youth regardless of whether they are present or not in your community at this time. Use the numbers preceding each of the values below in your ratings:

1—Extremely valuable; every student should participate.
2—Considerable value; extremely valuable to some students
3—Some value; a few students would find it worth while.
4—Practically no value; should not be offered.

ACTIVITIES	I	II	III	IV	V
Interscholastic					
Football					
Basketball					
Baseball					
Track					
Wrestling					
Music					
Marching band					
Concert band					
Orchestra					
Dance band					
Small ensemble					
Social clubs					
(Others may be continued in similar manner)					

FIGURE 10. Example of Structured Items.

interpretation than that which I have intended? If the answers to these questions are not satisfactory, it is likely that the questionnaire, even if returned, will yield scanty and unreliable information.

During recent years there has been a widespread use of "objective-type" tests, and extensive efforts have been made to improve test items. Since there is no sharp line of distinction between the form of a questionnaire and a test, the general rules for test construction may serve as guides in constructing a questionnaire. The following list of 16 rules for constructing a research questionnaire contains several that have also been presented in a discussion of test construction:[3]

1. Express the item as clearly as possible.
2. Choose words that have precise meanings wherever possible.
3. Avoid complex or awkward word arrangement.
4. Include all qualifications needed to provide a reasonable basis for response selection.
5. Avoid the inclusion of nonfunctional words in the item.
6. Avoid unessential specificity in the question or in the responses.
7. Avoid the inclusion of trivial questions.
8. Make the suggested answers as simple as possible.
9. Be sure the items will seem to the respondent to apply to the situation.
10. Refrain from asking questions of opinion unless opinion is what is specifically required.
11. Avoid items that are too suggestive or too unstimulating. They should not lead a respondent to go beyond the facts, but they should induce him to provide the required information.
12. Phrase questions to avoid the academically or socially acceptable responses. Make it possible for the respondent to answer truthfully without embarrassment.
13. Avoid questions that may be checked with several responses when only one response is desired. In fact, such items usually are quite difficult to tabulate and analyze.

[3] Robert L. Ebel, "Writing the Test Item," in *Educational Measurement*, (E. F. Lindquist, Ed.), Washington, D.C.: American Council on Education, 1951, pp. 213–216.

14. Whenever possible, questions should be worded in such a way that they can be answered simply by a check mark.
15. Ask questions in such a way that they will relieve the respondent of as much complex thinking as possible. A popular technique is to reduce a complex question to a series of questions which are easier to answer.
16. Avoid the use of words which are susceptible to different interpretations; e.g., moral or immoral, good or bad, rich or poor, intelligent or ignorant, laborer or capitalist.

Directions for Responding

In planning the questionnaire it is necessary to consider the directions for answering the questions, since to assume that *any* group of respondents can do without specific instructions is unrealistic. Because the respondent to a questionnaire cannot ask anyone for instructions—as he could if he were being observed or interviewed—the researcher must anticipate a respondent's need and introduce instructions essential to the understanding of specific items along with the items themselves. However, since space is limited, instructions should be as explicit as possible. Instructional examples are often helpful. Directions should be designed to promote uniformly accurate replies. They should stand out sharply by the use of boldface type, capital letters, italics, or underscoring. The set of instructions illustrated in Figure 10 exemplifies good techniques.

It is often difficult to make certain that the respondent has answered a question truthfully. There are many reasons why he may not do so: (1) he may not know the answer and be merely guessing, (2) he may not be thinking critically, (3) he may not have understood the directions correctly, (4) he may be apprehensive about telling the truth, or (5) he may feel that the question is too personal in nature. He should be assured, in the directions or in the covering letter, that his identity will not be revealed and that the information he provides will not be used in a manner adverse to his interests. Naturally, he should be impressed with the importance of answering the questionnaire completely and to the best of his ability.

Questions to Establish Favorable Respondent Rapport

The questionnaire method requires a degree of cooperation on the part of the respondent which is more difficult to obtain than in the observation or interview methods. In the mailed questionnaire method, rapport is established *only* through correspondence. The use of questionnaires has been so seriously abused by many researchers that cooperation now is sometimes difficult to obtain. Many times researchers have lost their perspective concerning what is reasonable to ask of another person who is usually a stranger to them. They have requested information on problems that have had little *apparent* value to anyone. It should be a rule *not* to send out questionnaires unless they deal with an obviously important problem.

The questionnaire should be so prepared as not to make any more demands on the time of the respondent than is absolutely necessary. It should be so presented that the respondent can see the value of the study and understand the investigator's point of view. In an attempt to establish favorable rapport, it is sometimes desirable to include some items for which information is not absolutely necessary—items whose main function is to put the respondent in a proper frame of mind. Three types of such items may be included at the beginning of the questionnaire for the following purposes.

To Get the Respondent's Mind on the Topic. This is sort of a "warm-up" type of item pertaining to some aspect of the problem that should be entirely neutral with respect to the respondent's emotions. It should be directed toward obtaining such information as type of school with which the respondent is associated, kinds of curriculum offerings, or number of teachers employed. These bits of information, while generally available from other sources, can usually be supplied by a respondent without much thought and probably with no reluctance. Questions dealing with more detailed information and data that may be considered somewhat private in nature should always be placed much later in the questionnaire.

To Allow the Respondent to "Let Off Steam." In some

cases it may be desirable to include some items which allow the respondent to express strong personal feelings as these may put him in a better mood to provide the information sought by the rest of the questionnaire. These are often open-end questions on which the respondent writes freely (unstructured items). There is some question, however, as to whether this technique is satisfactory or whether it creates additional emotional reactions that can invalidate the entire questionnaire. More research on this topic is needed.

To Avoid the Respondent's Opinion That the Questionnaire Is Inadequate. As a rule it is not desirable to ask a respondent for information that can be obtained elsewhere easily and with complete accuracy, but in some cases a few such questions may be included because they belong "logically" to the composite of items for which information is sought. They are included only for the purpose of eliciting responses to the entire questionnaire. For example, a questionnaire soliciting the requirements for teachers' certificates that did not include questions with respect to the total number of credit or semester hours of instruction required is likely to raise questions in a respondent's mind as to the competency of the researcher. Questionnaires that do not appear to cover a topic adequately are often disregarded by respondents.

Length of the Questionnaire

As a general rule, a questionnaire should be long enough to include all the information essential to the study, but not so long that the respondent will reject it as being too time-consuming. A short questionnaire usually stands a better chance of being answered than a long one. However, if it seems necessary to have a long questionnaire to secure adequate information upon which to base valid conclusions, it should be developed to the length needed even though the percentage of returns is likely to be small, especially as this can be offset somewhat by sending the instrument to a larger sample of the population being studied. The length of the questionnaire should

be dependent *entirely* upon the extensiveness of the data required and should not be controlled by the expected number of returns.

Long questionnaires may often be made to appear shorter than they are by the form in which they are produced. Those that are typewritten or mimeographed usually appear to be much longer than those that are printed. Very attractive instruments can be printed in small type so that one page of the questionnaire will include the same amount of material as three or four typewritten pages. In some cases it may be possible to make the instrument appear shorter by grouping items into several sections and numbering each section. Using this system the last number on the last page will not be numerically as large as it would have been had the questionnaire not been divided into sections. Also, questionnaires that are printed with small but easily legible type and that have ample spacing between items—considerable blank spaces showing—will not appear as long as those printed with larger type where the items are crowded on a page.

Design to Facilitate Tallying Procedures

As soon as the questionnaires start to come back to the researcher, he will want to begin tallying the responses. Thus, he should keep in mind some procedure to facilitate the tallying whether he will be doing it by hand or using a data-processing machine. If the answers to items can be arranged along the right-hand margin of the questionnaire, it is easier to locate and tabulate them than if they are scattered throughout the pages. For data-processing machine tabulation it is desirable to have all items precoded as this facilitates transcribing them to machine punch-cards. Details of this process are discussed more fully in Chapter X.

Sometimes questionnaires with several pages may be offset at the time they are stapled together so that the margins of all pages can be seen without turning pages. This simplifies the tallying procedures considerably. Figures 11A and 11B illus-

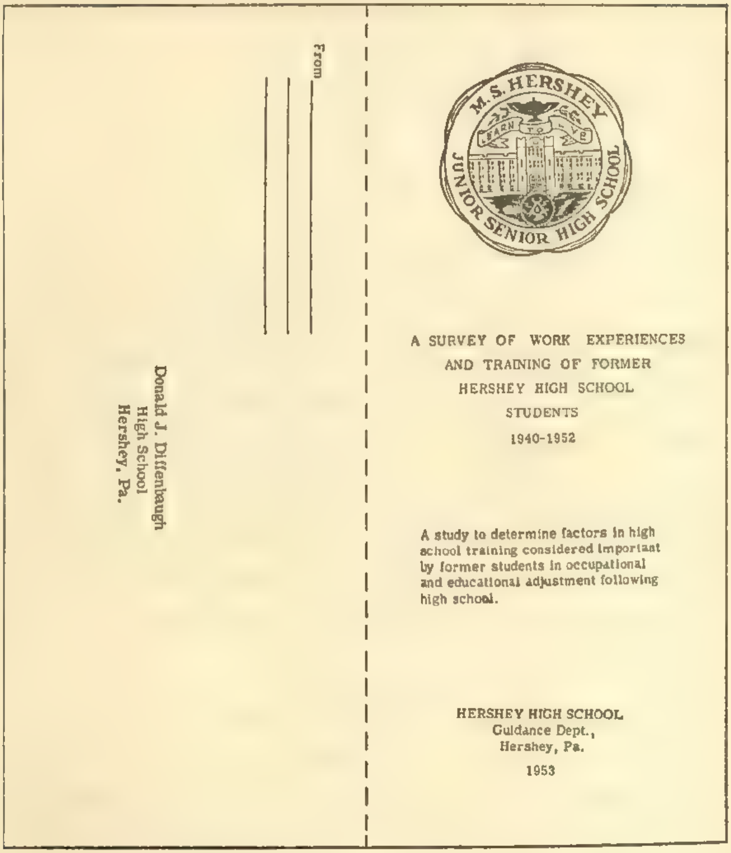


FIGURE 11A. Ideal Format for Front and Back Covers of a Questionnaire Booklet.

trate a questionnaire used in a follow-up study of high-school graduates to determine the adequacy of their education for their present type of work.

Figure 11A shows the front and back of a questionnaire booklet that incorporates practically all the desirable features of questionnaire format. Figure 11B shows the inside of the front cover, the first page of the questionnaire, and the edges of the following three pages. The objective data that could be easily tabulated were placed on the offset pages. On the backs of

Dear Former Student:

This is a check-list you said you would complete and return to us. The members of the faculty recognize the need to secure information from former students regarding the value of their high school training. your answers can be used as a basis for making high school more valuable to students in the future.

You will help us to find answers to the following questions:

1. What jobs are former students now doing or planning to do?

2. What factors do former students consider important in getting along well in a job?

3. What factors in high school training were helpful in making vocational and educational adjustments after leaving school?

4. What phases of high school training can be made more effective in helping students in the future to better job adjustment?

A summary of the results of the study will be published in our school newspaper, "THE BROADCASTER" in the late spring of 1953. A copy will be sent to you upon request.

Your name need not be attached to this check-list unless you so desire. All information will be kept confidential.

Thank you for your help and cooperation.

Sincerely yours,

Donald J. Diffenbaugh
Guidance Committee
Chmn., High School

I. EMPLOYMENT DATA

A. If you are not engaged in work in the field for which you planned in high school, check reasons below:

1. Not a good choice because of:

a. Lack of knowledge of my abilities and interests ☐

b. Too young to make a wise choice. ☐

c. Lack of knowledge of job requirements ☐

d. Parents' objection ☐

e. Other: ________________ ☐

2. Little opportunity in this field ☐

3. Needed more training and couldn't get it. ☐

4. Other: ________________ ☐

B. Annual Income (1952). If both man and wife are working, check family income on right.

Your earned income on left.

Above $7000	8	☐	☐
$6000—6999	7	☐	☐
$5000—5999	6	☐	☐
$4000—4999	5	☐	☐
$3000—3999	4	☐	☐
$2000—2999	3	☐	☐
$1000—1999	2	☐	☐
Below $1000	1	☐	☐

FIGURE 11B. Efficient Organization to Facilitate Tabulation of Questionnaire Items.

these pages were included items that required short answers and free responses to specific questions. Questionnaires of such exceptionally high quality, not only in appearance, but in content as well, should be the goal toward which graduate students should work.

QUESTIONS TO TEST VALIDITY OF RESPONSES

Many times respondents fail to recall information as accurately as they should, or they unconsciously or deliberately

exaggerate or grossly underestimate. Therefore the researcher should include some questions in his questionnaire that can be used to check on the validity of other items.

Some discrepancies may be identified by checking independent sources—birth records, work certificates, payrolls—but this process can be extremely time consuming and costly, and often such records are not available to the researcher. It may be desirable, instead, to include items which solicit nearly the same information at widely separated places in the questionnaire or to ask several questions to identify contradictory or implausible answers. For example, one question might elicit the response that a mother was 39 years of age; another that her oldest child was 26; another that she was graduated from high school in 1928; and another that she had been married 27 years. On the basis of all these replies the researcher might well assume that the mother was considerably older than she had stated because the latter three statements, which tend to corroborate each other, would indicate her age closer to 50 than 39.

TECHNIQUES TO FACILITATE USABLE RETURNS

Since a questionnaire mailed to a number of informants is merely "several pieces of paper" which lack any of the personal charms or social skills of the researcher, it must be presented in a manner that will attract the recipient's attention and induce him to answer and return it. The basic technique for soliciting responses is to send a covering letter which should be brief but should explain what the researcher is doing, why he is doing it, and who is sponsoring his study. Other means of facilitating a high percentage of responses are: guaranteeing the respondent that his answers will be held in confidence, enclosing stamped return envelopes or participant reply cards, using various follow-up procedures, selecting an appropriate time to send the questionnaire, and making certain that the instruments used have a professional appearance.

The Covering Letter

The covering letter should explain the nature of the research

problem, its importance, and the necessity of having answers from the recipient. It should attempt to arouse the recipient's interest so that he will complete the questionnaire fully and truthfully and return it within a reasonable time. Some of the essential components of the covering letter are cited here with a brief discussion of each.

Appeal Through the Purpose of the Study. The purpose of the study should be stated frankly and concisely to allay suspicion on the part of the respondent concerning any hidden or ulterior purpose. It need not be elaborate, but it should be sufficient to explain why the researcher needs answers to all the questions asked. The respondent should be pursuaded to participate by arousing his interest in the purposes of the investigation rather than by appealing to his self-interest or by pleas to help the researcher. Goode and Hatt believe that the most effective appeal is an altruistic one.

> Although these special inducements such as money have increased the proportions of returns slightly, it remains true that no advantage that a research organization is likely to be able to offer will appeal to a large group of respondents. The amount of money that can be offered is trivial, and so are the other advantages, compared to the amount of time and thought requested. Whatever the student may believe concerning the cynicism of the age or the selfishness of people, extensive research has demonstrated that an appeal to disinterested motives is strongest. "The information is needed by thousands of leaders attempting to solve today's problems," "You will be contributing to the advancement of science," and "You will help improve the education of thousands of students who will attend Xiphosuran College in the future" are all better arguments than attempting to exploit the self-interest of the respondent.[4]

If it is the case, it may be desirable to let each respondent know that he is one of a select list who are being asked to help. If the person to whom the questionnaire is sent is one member of a scientifically chosen random sample, the importance of his answer and the difficulty of substitution may be stressed

[4] William J. Goode and Paul K. Hatt, *Methods in Social Research.* New York: McGraw-Hill, 1952, pp. 177–178.

because many people are now sophisticated about sampling procedures and realize that their failure to cooperate could be a serious blow to the study. This should never be done, however, if the statement is untrue. Aside from any ethical implications, an untruth, if found out, would tend to prejudice further attempts to secure cooperation. Figure 12 gives a good illustration of a "problem-centered and partner-centered approach" in a covering letter. The letter describes the problem

Dear Professor,

Newspapers and radio in this country have received an increasing amount of criticism from a wide variety of sources. Usually, however, this criticism comes from individuals with axes to grind and often goes unheeded as a result.

The enclosed questionnaire is a part of a study being made to find out something of what a specialized group thinks of the press. The survey group consists of the 195 persons now on the University of ________ faculty who hold the academic title of professor, associate professor or assistant professor.

As a teacher you are one of a small but influential group of opinion leaders. Therefore, the opinions you hold of the newspaper and radio press are important. It is believed that the press and the consuming public will be benefited by examining responses to studies of this kind. The press will be served -- and may serve better -- by its knowledge of what opinion leaders think of it. Such a study will give some indication of the degree of agreement this group holds with charges made by individual critics. It would be of value to know, for instance, what opinion leaders think of governmental control of the press, or how effectively the media influence public opinion.

The questionnaire is to be strictly anonymous. For the purposes of my study, I am interested in totals, not in names. The value of this study will be greatly increased if respondents provide:

1. a candid answer to every statement or question;
2. prompt consideration.

I would appreciate your giving the questionnaire your considered judgment and returning it to me by campus mail as soon as possible.

Sincerely,

FIGURE 12. A Problem-Centered and Partner-Centered Approach in a Covering Letter.

being investigated and suggests that the respondent is important and is one of a select group.

Beginning researchers often develop good research proposals as they know what information is necessary, but they fail to collect this information because they send out badly composed covering letters. The "ego-centered" letter illustrated in Figure 13 was accompanied by a well-constructed questionnaire but the

To Director of Student Teaching:

At the present time I am engaged in gathering information for a doctoral dissertation. The purpose of my study is to describe, compare, and analyze the major types of methods used to provide student teaching experiences at the secondary level. In order to determine what types of methods are most widely used it is necessary to make a survey of teacher training institutions.

You and your staff can be of great assistance to me by filling out the enclosed questionnaire and by returning it at your earliest convenience. Please feel free to make any additional comments or remarks that you may wish to make. I am most desirous of obtaining as complete a picture of each program of student teaching as possible.

I realize the questionnaire may seem lengthy, however, everything included is felt to be important. It has also been constructed so that it may be checked quite easily. Depending upon your program it may be possible to omit Part II or Part III.

Thank you for your help in this matter. Your time and contribution is truly appreciated.

Sincerely,

Enc.

FIGURE 13. An Ego-Centered Covering Letter.

researcher would have been quite fortunate indeed to have received adequate returns. The researcher appeals for information for himself—note the numerous mentions of "I" and "to me"—rather than arousing the respondent's interest by stressing the importance of the problem. The "no-purpose" request, shown in Figure 14, is even worse and to expect returns would be entirely unrealistic. Besides containing a misspelled word, there is no indication at all of the researcher's problem. Both

questionnaires were received from different students working toward master's degrees at the same institution. One might infer that the quality of graduate work at that institution is questionable!

Dear Sir:

I am seeking information for a Master's thesis. I will greatly appreciate your co-operation in filling out the attached card and returnung it at your earliest convenience. Thank you for your consideration.

Sincerely,

Daily practice	Time start	Length of practice
Basketball	____	____
Baseball	____	____
Track	____	____

Time games start	Basketball
B game --------	____
A game --------	____

FIGURE 14. A "No-Purpose" Postal Card Questionnaire.

Auspices and Endorsements. The more important the recipient of a questionnaire considers the research topic to be and the more impressive the support for the study, the greater are the possibilities he will give a good response. Just as setting up an interview may be facilitated by using intermediaries, so the completion of a questionnaire may be facilitated by naming those who are supporting the study. If permission can be obtained, a covering letter should be sent out on the letterhead of the institution in which the study is being pursued. In many

cases it is better if such a letter carries not only the signature of the researcher as "Project Director," but also the signature of a faculty member, department head, or administrative officer of the institution. In addition, it is sometimes desirable to have a separate endorsement, either at the bottom of the covering letter or in an accompanying letter, written by some individual or group which has the respect of the recipient, such as an officer of an educational institution or some trade or professional organization closely related to the study problem. In some cases, the covering letter may devote a few lines to explaining the character of the organization or agencies endorsing the study. Figure 15 illustrates a covering letter in which the researcher acknowledges that his questionnaire will make demands on the time of the respondent, points out the importance of the study to "those of us who are classroom teachers," states the educational agencies endorsing the study, assures the respondent that his confidence will be kept, and includes an endorsement for intermediary support.

An excellent example of covering letters and questionnaires in which an intermediary is used to improve the questionnaire return is illustrated in Figures 16A–16D. The questionnaire, with a separate printed sheet describing the study, the covering letter, a participant reply card, and a letter by an intermediary were all enclosed in the same envelope. The intermediary's letter (Figure 16A) was typed on university letterhead stationery and addressed personally to the respondent urging cooperation on the basis of the respondent's professional work. The covering letter, Figure 16B, was also mimeographed and addressed personally to the respondent. The title of the study and the address where it was being conducted was typed at the top of the page. This letter calls attention to the separate enclosed description of the study, Figure 16C, appeals to the respondent as an authority in the area of investigation, describes the need for the study, and asks the respondent to return the participant reply card if he is willing to cooperate with the study. Figure 16C shows the printed description of the study. This is very professional in appearance and provides for a more detailed

Eugene, Oregon
October 15, 1951

To The Classroom Teacher:

Although there are numerous demands on your time, will you take a few minutes for a task which may have significance in the improvement of teaching positions such as yours?

Those of us who are classroom teachers realize the importance of teacher morale and efficiency in a school system. The writer is conducting "An Investigation of Elementary and Secondary Teacher Beliefs and Opinions About Their Working Conditions in the State of Oregon." This study is being made under the direction of the School of Education of the University of Oregon. It has been endorsed by the State Department of Public Instruction and by the Department of Classroom Teachers of the Oregon Education Association.

By obtaining answers from a large number of classroom teachers to the questions submitted on the attached questionnaire valuable information should be provided to improve the working conditions of teachers.

Will you cooperate in this investigation by completing the attached questionnaire at your earliest convenience and returning it in the business reply envelope provided for your use? In return for your consideration of this questionnaire the results will be made available to the teachers of Oregon.

No teacher, school, or school system will be identified in the results of the study. It is not necessary to sign your name.

B. Willard DeShazo.

"I urge cooperation of teachers with B. Willard De Shazo in study *Teacher Beliefs and Opinions About Their Working Conditions.*"

REX PUTNAM
State Superintendent of Public Instruction

FIGURE 15. A Psychologically Sound Covering Letter.

Professor ________
School of Education
University of ________
City, State

Dear Professor ________:

In the enclosed communication, Mr. LS is seeking your cooperation in a study being conducted by him and under my advisement. Your cooperation as a science educator will be of great value to his study, and I join with him in the hope that you will find it possible to participate.

Cordially,

FIGURE 16A. Intermediary Endorsement Accompanying a Covering Letter and Questionnaire.

Professor ________
School of Education
University of ________
City, State

Dear Professor ________:

Enclosed is a brief description of a study that I am conducting at the University of ________ in Science Education.

The group of judges referred to in the description is to be drawn from the population of science educators, examiners and administrators, and members of the educational foundations fields. The rationale for such selection is, probably, self-evident to you.

We who are in the field of science teaching realize the need for proper training of the young people who are fitted for, and interested in becoming science teachers. Since your training and interest is in science education, your cooperation in this phase of the study is needed, and I am, therefore, asking your assistance. The findings of this study will be published and sent to each person participating in this research.

Will you return the enclosed reply card indicating whether or not you are willing to complete the competence rating form. If you reply affirmatively, that form will be sent to you for your attention.

Sincerely yours,

FIGURE 16B. A Well-Prepared Covering Letter.

A Study In Science Teacher Competence

During the past six years several studies have been conducted at Stanford University in the area of teacher competence. The initial planning and edification was a committee effort and subsequent, more definitive efforts have rested with doctoral research. The first of these studies dealt with teaching competence in general; more recent studies have been concerned with specific aspects of competence and specific teaching fields. This material is now being used by the National Commission on Teacher Education for Professional Standards as a basis for stimulating thinking on a national scale.

The present study is an attempt to define the competences necessary for the secondary school science teacher. It is hoped that such a study will help the teacher training institutions in their task of preparing secondary science teachers.

An initial list of competences has been selected from related studies and from publications in the field of science education. A group of presumably competent judges are being asked to assist in this study by rating the importance of the competences, as they see them in light of their experiences. Further, they are being asked to indicate which competences should be emphasized in pre-service training, and which may be deferred to in-service training.

The form that is being sent to the judges, required only the rating of the items of competence and the checking of the place of emphasis as outlined above. No other information will be solicited. In preliminary try outs of this task judges were able to complete the form in thirty minutes or less.

FIGURE 16C. Independent Problem Statement Accompanying Covering Letters and Questionnaires.

description than it would be desirable to include in the covering letter without making it unduly long. The first page of the long questionnaire is presented in Figure 16D. It gives detailed directions for completing the questionnaire and explains the specific meanings of several of the terms used. The questionnaire itself was constructed with the questions on one half of a page, the response spaces on the other half, and perforations

down the middle. When the questions were all answered, the respondent detached the answer part of the questionnaire and mailed it in the envelope provided. While this was a time-consuming questionnaire, the manner in which it was constructed and sent out merited consideration by the persons to whom it was mailed.

Guarantee of Anonymity or Confidentialness of Respondents. There seems to be a prevalent belief among some research workers that better and more complete returns will be forthcoming if a respondent does not have to identify himself. Research in this respect, however, is meager and inconclusive. It is conceivable that very personal information or data that may be embarrassing to the respondent would be more easily obtained if the respondent knew he would remain anonymous. On the other hand, if the respondent is asked to give information because he is one of a select group whose personal judgments are considered important, it is conceivable that he will not only be willing, but honored, to have his name attached to his questionnaire return. Some recipients may feel that the omission of the signature detracts from the importance of the study, or that if the questionnaire is worthy of their careful efforts, it is worthy of their name, or that the omission of the name and address implies no follow-up and hence rather poor standards of data collection.

If a questionnaire is to be *truly anonymous,* the covering letter must include a guarantee that the respondent will remain anonymous; there can be no request for names and addresses and no questions which are so detailed as to make identification easy. Thus, the researcher will not be able to describe his sample, and his conclusions can have no value without an adequate description of the sample or the population it represents. In addition, the use of data collected anonymously limits the researcher to making only rather insignificant and cursory studies.

In those instances in which the recipient's name is not important, it may be desirable to have him respond with an *anonymous attitude;* that is, he should feel perfectly free to say

SCIENCE TEACHER COMPETENCES AND THE FUNCTIONS OF SCIENCE IN THE SECONDARY SCHOOL

The institutions preparing science teachers need your assistance in helping them determine the following: 1) What are the most important responsibilities of secondary teachers of science? 2) What competences should be emphasized in the pre-service training program? 3) What competences can be deferred to the inservice training program?

Professional teaching competences, not specific subject matter competences, are those under consideration.

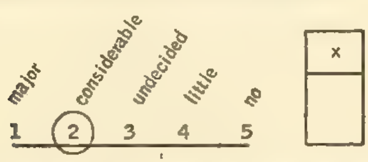
DEFINITION OF TERMS. Competence is the knowledge and understanding, ability and skill, attitude and philosphy which should be brought into action by the teacher of secondary science. Function is the obligation or duty of science in the secondary school. Secondary School includes grades nine through twelve. Science Teachers includes all full-time teachers in the fields of life or physical science.

The functions and competences have been selected from professsional literature. To make them more useful you are asked to act as a judge in estimating the relative importance of the functions of science in the secondary school and the competences needed, as you see them, in their relationship to the field of science teaching.

DIRECTIONS. Rate each function and competence as you see its importance. The rating scale is designed to read as follows:

1 indicates major importance
2 indicates considerable importance
3 (midpoint) undecided as to importance
4 indicates little importance
5 indicates no importance

Encircle the number of your choice as illustrated.



The column headed x to the right of the rating scale may be checked in case the meaning of any statement is not clear.

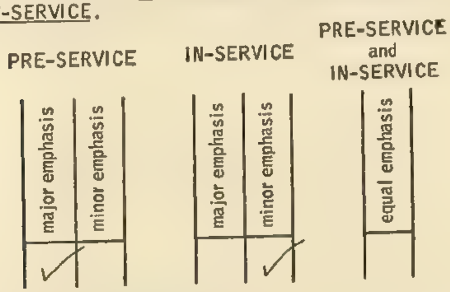
The columns to the right of the rating scale are provided for you to indicate, on the basis of your experience, where you believe the emphasis in the training program should be given for each competence which you have rated 1 or 2. Note: This part pertains only to the competences, not to the functions.

FIGURE 16D. Clear-Cut Instructions for Questionnaire Responses.

If you believe that the competence should receive its entire emphasis in the pre-service training program place a check in the major emphasis column under PRE-SERVICE. However, if you believe major emphasis should be given during the pre-service training program but that some emphasis should be given during in-service training, then also place a check in the minor emphasis column under IN-SERVICE, in addition to the mark in the major emphasis column under PRE-SERVICE. This case is illustrated below.

If you believe the competence may be deferred to the in-service training program place a check in the major emphasis column under IN-SERVICE. However, if you believe that major emphasis may be deferred until in-service training but that some emphasis should be given during the pre-service training program, then also place a check in the minor emphasis column under PRE-SERVICE in addition to the mark in the major emphasis column under IN-SERVICE. The marking in this case would be the reverse of the one illustrated below.

If your experience leads you to believe that equal emphasis should be given to the competence in both the pre-service and the in-service training programs, then place a check in th the column marked equal emphasis under PRE-SERVICE and IN-SERVICE.



Timing by judges in marking this form in preliminary try outs did not exceed thirty minutes.

DIRECTIONS FOR RETURNING. When you have completed checking this form, please detach the answer sheets along perforations and return in the provided envelope.

Your cooperation is appreciated. Thank You.

FIGURE 16D. *(Continued)*

whatever he wishes without fear that his responses may be used in a manner contrary to his interests. In these cases the covering letter should never state directly that he cannot be identified, and it should not make any use of the word "anonymous." It is much better simply to state that he does not need to sign his name (see Figure 15).

If the questionnaire was sent to a probability sample of the population, the researcher may wish to know what questionnaires are outstanding so that he can either trace them or make substitutions. If the sample is stratified, it is essential to have some idea of the questionnaires that are outstanding or have been received. Several methods have been proposed for identifying questionnaires in which respondents have provided information assumed to be anonymous or in which confidence has been guaranteed. One excellent procedure is described in connection with a "capital earnings" research study reported by Taussig and Barker. The covering letter sent out included the following description of their procedure to protect the confidence of their respondents:

> This form when sent you bears no identifying number or other mark. After you have filled it in, sign the card, put it in the small envelope and seal it. Both the return and the card may then be mailed in the large envelope at any post office you please. Thus the return will come to us without any means of identification except the card within the sealed envelope, addressed and assigned to Professor F. W. Taussig exclusively. Upon receipt we shall give the return and the sealed card a number in duplicate. Thereafter your report shall be known only by this number.
>
> In all scientific investigations there must be a means of identifying all data. This is our sole reason for requiring the signed cards. They will be held under seal by Professor Taussig; he will be personally responsible for them. Further, as to this return, though there will be no means of identification whatever upon it, no one outside this department shall see it. You can in no way be embarrassed through disclosure of information.
>
> We realize we have asked you to undertake a prolonged and somewhat intricate examination of your records, yet we believe the results

will amply justify the work required of you and of us. You will join in a constructive way in the advancement of scientific business.[5]

Another procedure is to give a frank statement in the letter of transmittal such as the following:

> The identity of the individual teacher or school executive is not sought, and if such identity is revealed to us, it will be held in *strict confidence.*

By using such a statement as the above it is possible for the researcher to identify each questionnaire return and at the same time permit each respondent to feel that he has given the required information anonymously.

Indirect or concealed identification may be accomplished by coding each questionnaire in various kinds of invisible ink, although this is not desirable since the ink may be detected. A more satisfactory, and almost undetectable method, is the following:

1. Type a list of all individuals in your sample to whom questionnaires are to be sent along the sides of a sheet of paper the same size as the questionnaire and place a period at the edge of the sheet by each name as shown in Figure 17.
2. Place a questionnaire alongside this sheet and make a pinprick near the edge of the questionnaire corresponding to the period at the edge of the identification sheet for the individual to whom the questionnaire is to be sent.
3. When the questionnaires are returned they can be matched up with the identification sheet and numbered according to the number assigned to each person listed.

Names can be typed along all four borders of the identification sheet. Several identification sheets may be used if the number of the sample is large and the pinpricks on the corresponding questionnaires may be made right at the edge of the

[5] F. W. Taussig and W. S. Barker, "American Corporations and Their Executives: A Statistical Inquiry," *The Quarterly Journal of Economics,* November, 1925, p. 1–51.

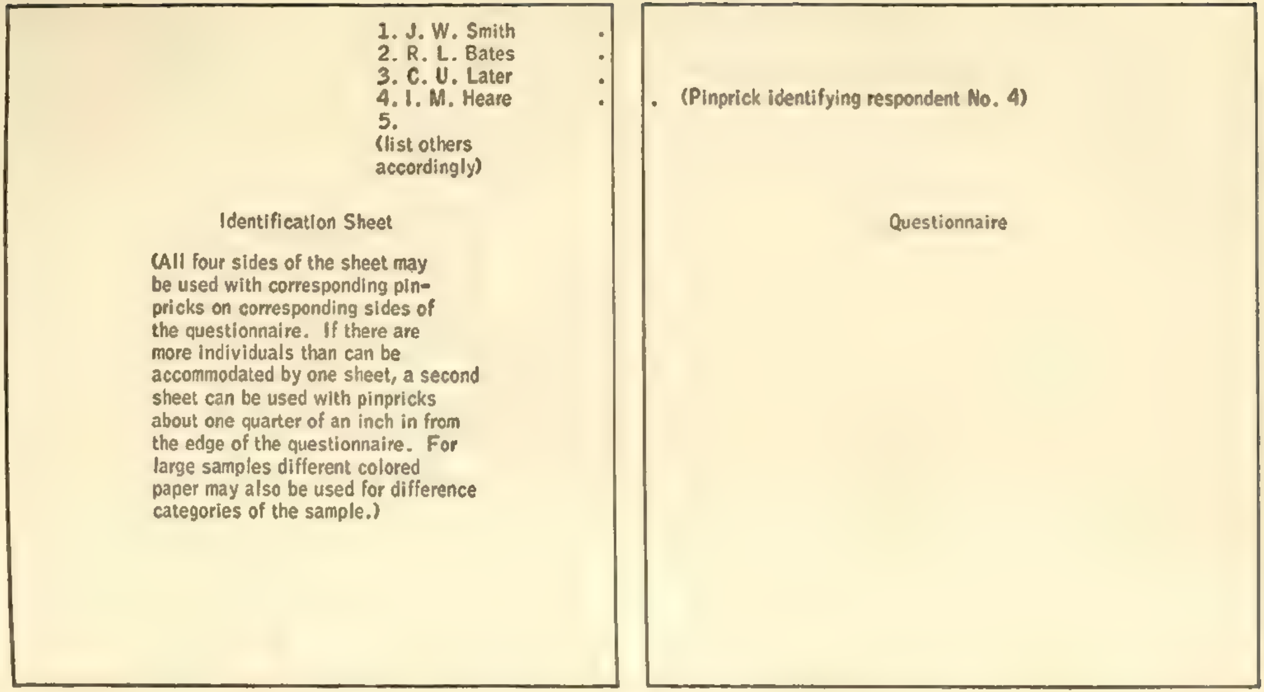


FIGURE 17. Method for Identifying Questionnaire Returns.

first identification sheet, a quarter of an inch in from the edge for the second identification sheet, and so on as required. This method is practically impossible to detect and, if detected, it is very difficult to disturb even if the respondent makes additional pinpricks because the chances of these pinpricks lining up with the periods on the identification sheet are very slight. Even if one respondent made additional pinpricks which matched up with a second respondent's name on the identification sheet, it is improbable that the second respondent would put pinpricks in a position which matched up with the first respondent's name on the identification sheet. Thus, identification could still be established. In using this method, anonymity is implied and *it is the moral responsibility of the researcher to see to it that the confidence of each respondent is not broken.*

Use of Return Envelopes and Postage

There is lack of agreement regarding the desirability of providing return envelopes and postage when sending out questionnaires. Some research authorities contend that the researcher cannot seriously expect the respondent to exert himself to the extent of providing his own envelopes and postage for returning a questionnaire. Others believe that if a respondent is interested enough to spend the time and effort to fill out a questionnaire conscientiously, he will not hesitate to return it even at his own expense. Furthermore, questionnaires sent to business firms when completed are usually placed in an outgoing mail basket on an executive's desk or are handled in a routine manner by a secretary. In either case the presence of a stamped return envelope will have a negligible effect, but the cost to the researcher for return mailing of all questionnaires may be considerable. On the other hand, a higher proportion of questionnaires sent to individuals will probably be returned if a stamped return envelope is enclosed. If no envelope is enclosed, the address of the person or agency to which the completed questionnaire is to be returned must be placed so that it is very easy to find.

Some research authorities hold that when questionnaires are sent out better results are obtained by using envelopes affixed with regular stamps instead of the stamped envelopes from the post office or those stamped by a postage meter. They believe using regular stamps avoids giving the impression that the questionnaire is associated with an advertising or selling campaign and lessens the likelihood of the questionnaires being thrown away.

Use of a Participant Reply Card

Sometimes it is desirable to include with the covering letter and questionnaire a return postal card to be mailed back immediately acknowledging receipt of the questionnaire. This card can be printed with blanks for the respondent to give his name, title, address, phone number, or other identification data. It may also include identification information of the person responsible for answering the questionnaire if he differs from the one to whom it was sent originally. This will enable the researcher to determine the extent of coverage he can expect and will provide a basis for follow-up reminders if questionnaires are not returned soon after a given deadline.

In some instances the postal card reply form is sent with a letter requesting participation in a study some time ahead of sending out the questionnaires. This may be done to avoid the expense of sending out bulky questionnaires to anyone who has not indicated he will participate in the study, or it may be done to determine whether the selected sample will yield sufficient returns for a valid study or whether an additional number of persons need be added to the original sample. The card shown in Figure 18 was sent out prior to the questionnaire illustrated in Figure 11 (p. 130) to enlist the cooperation of respondents and included some items for purposes of precategorizing the information later obtained by the questionnaire.

Follow-Up Procedures

Unless a researcher uses some type of follow-up techniques to solicit responses, he is often likely to receive an insufficient re-

Derry Township Schools
Hershey, Pa.

Date ----------

Dear Former Student:

It may seem to you that high school days are deep in the past. Actually, there are many remaining evidences of your contribution to Hershey High School.

Are you willing to contribute further by completing a check-list concerning your job and educational experiences since high school? Your answers will help us to give the students better services. If so, kindly fill in the attached card and send to the addressee. The check-list will then be sent you which will require a few minutes time to complete. Your name need not appear on it. All material will be kept confidential.

Sincerely yours,

Donald J. Diffenbaugh, Chmn.
High School Guidance Committee

No. ---- Class of
Name ---------------------------------- 19 ---- or Gr. ----
(Include Married Name)
Address --
No. Street City State
When in H. S., I lived in ------------------------ School District
Course in H. S. ---------------------------------- Sex. M ☐ F ☐
When in H. S. I selected ---------------------------- as a career.
At present I am a 1 Student ☐ 2 Housewife ☐ 3 Self Employed ☐ 4 In Armed Services ☐ 5 Employee ☐
1 Married ☐ 4 Widow(er) ☐ No. of Children
2 Single ☐ 5 Divorced ☐ ---- Boys Ages ----
3 Separated ☐ 6 Other ------ ----Girls Ages ----
Present job --
Did you secure schooling beyond H. S.? Yes ☐ No ☐
I am willing to help in the follow-up study by completing a check list.
Yes ☐ No ☐
Date ------------------

FIGURE 18. Participator Solicitation Card with Predetermined Classification Information Requested.

turn of the completed questionnaires. Various techniques have been suggested, among which the following have appeared to be the more promising.

1. About two weeks after sending out a questionnaire, mail a card or letter to the nonresponding recipients calling their attention to it. If necessary send out a second reminder.

2. It is sometimes helpful to make a second mailing of the questionnaire with a new cover page or new accompanying letter stating the possibility that the original questionnaire may have been misplaced or "buried" in a mass of correspondence.
3. A personal letter, individually written and signed, may be enclosed with a second questionnaire making a special appeal for cooperation. At this time a stamped return envelope is certainly indicated as it may entice a response.
4. One of the best procedures is to telephone the nonrespondent. This may incur long distance bills, but the researcher can usually learn whether or not he may expect cooperation.
5. A short form of the questionnaire may be sent by airmail, special delivery, or registered mail. This form should include the more essential items. When he receives back the short form, the researcher may then send out to those who responded to the short form a supplementary form including the other required items. A person who has replied to the first form is quite likely to provide the additional information on the supplementary form. (This technique is sometimes used instead of attempting to obtain responses from one mailing when the questionnaire seems unduly lengthy.)
6. In some types of studies the researcher may partially fill out the questionnaire for the respondent if several of the items include information readily available from other valid sources. This is then sent to the respondent, and he is asked to check the filled-in items for accuracy, and at the same time an appeal is made for him to provide other needed information.

Timeliness

Questionnaires should be sent out to arrive when they have the best chance for consideration. The researcher should avoid periods that are known to be especially busy ones for the recipients. The timeliness of the study is important since it will arouse greater interest if it is carried on near the time the phenomena being investigated occurs. For example, studies of the effects of an election or of recently enacted legislation upon certain businesses will probably receive more favorable attention near the time these effects are first felt than at a much later date.

Professional Appearance of the Questionnaire

Cover letters and questionnaires with a professional appearance are more likely to be completed and returned. To obtain this appearance the researcher should consider typography, spacing, choice of paper, and method of reproduction. Attractive typography—print if possible—a good grade of paper, good spacing, and a generally neat appearance create a favorable impression on the recipient. They indicate that the researcher has considered his investigation important enough to warrant the additional time and expense necessary to make his data-collection instruments attractive.

Figures 19A and 19B illustrate the covering letter and the first page of a questionnaire used in a study conducted by a university professor. While the study was very important to education and was well-designed, the format of the questionnaire and covering letter was poorly developed. The covering letter, Figure 19A, was mimeographed on university letterhead stationery. While the content of the letter was satisfactory, its appearance was unattractive. The questionnaire, Figure 19B, was also mimeographed and its contents were very crowded on each page. In terms of the importance of the study and the number of professional persons to whom it was sent, it deserved to be printed in an attractive form. Good printing and spacing of items, with plenty of blank space showing, would have made this a good questionnaire and covering letter instead of the less than satisfactory ones they are in their present form.

The business reply card may be used at times to get considerable information from respondents. Figures 20A and B illustrate a questionnaire for which the responses are returned on a business reply card. This questionnaire was mailed out with an accompanying letter describing the study in greater detail and making an appeal for cooperation based upon the respondent's desire to improve his own profession. In addition, the questionnaire includes a brief descriptive title, sponsorship, purpose, and directions to the respondent. The mechanical

Head of the Department of Education
University of ______
City, State

Dear Sir:

During the coming months I plan to work on the topic, "The Importance of Certain Characteristics of Teachers of Education in Programs Leading to a Baccalaureate Degree in A.A.C.T.E. Member Colleges and the Degree to Which These Characteristics Appear to Exist," for a doctoral dissertation at the ______ State College. This project has been developed with the approval and advice of Dean ______ of ______ State College, who also heads an A.A.C.T.E. evaluation committee.

As a result of the study, it is hoped that real strides toward the following objectives may be achieved:

1. To determine what characteristics are considered most important and how frequently they appear to exist.
2. To attempt to discover the relative degree of importance of these characteristics.
3. To compare attitudes of various groups of educators on this topic.
4. To enable educators better to understand each other in regard to teacher qualifications.
5. To help teachers of Education toward better self-analysis.

A questionnaire has been developed as an aid in accomplishing the above purposes and is to be used with the following groups: college presidents, heads of departments of Education, teachers of Education, students of Education, employers and supervisors of public school teachers, and public school teachers. However, before the questionnaire is to be given wide circulation, it seem important that it be reviewed by a number of persons in each group who are willing to serve as "jury members."

I should feel honored indeed if you would consent to look over the questionnaire and indicate your impressions on the enclosed blank. It is not necessary that you answer the questionnaire; your impressions alone would be most helpful.

Let me take this opportunity of expressing my deep appreciation for any help you can give me in this matter.

Sincerely,

FIGURE 19A. A Good Covering Letter Spoiled by Unattractive Format.

arrangement shows considerable ingenuity on the part of the researcher. The card was sent out folded so that the questions shown in Figure 20A were lined up with the responses to the right of the black line on the business reply card as shown in

DESIRABLE CHARACTERISTICS OF COLLEGE TEACHERS OF EDUCATION

Instructions

In checking the traits and experiences listed below as characteristics of college teachers of Education, please indicate your judgment of the practical importance of each as it affects Education teachers as follows:

1. Check the first column if the item is unimportant or only slightly so.
2. Check the second column if the item is of moderate importance.
3. Check the third column if the item is extremely important.
4. Check the fourth column if the characteristic has NOT been typical of most of the Education teachers whom you have known. Do not check items in Column 4 or 5 unless your experiences have been such that you feel able to indicate an opinion.
5. Check the fifth column if the characteristic seems typical of most Education teachers whom you have known.

Special lines have been provided at the end of each section for the addition of other items which the reader feels should be included.

I. AS A TEACHER:	IMPORTANCE SLIGHT (1)	MODERATE (2)	EXTREME (3)	NOT TYPICAL (4)	TYPICAL (5)
a. Selects teaching methods in terms of objectives.					
b. Emphasizes practical application of educational principles studied.					
c. Uses methods designed to secure wide student participation.					
d. Has studied problems of college teaching (formally or informally).					
e. Uses new and varied methods and materials in instruction.					
f. Understands problems most commonly faced by students in their training and in their professional careers.					
g. Has had successful non-college teaching experience.					
h. Believes in "General Education" for students in Education.					
i. Leads students to think for themselves.					
j. Helps students to express their ideas efficiently and effectively.					
k. Has enthusiasm for teaching that inspires others to want to teach.					
l. Allows students an opportunity to help plan work to be done in courses.					
m. Organizes work so as to provide a maximum or real, rather than "text", experiences.					
n. Organizes and prepares work carefully.					
o. Maintains a democratic classroom atmosphere.					
p. Evaluates self as well as students.					
q. Makes definite day-to-day assignments.					
r. Pays careful attention to physical conditions in the classroom.					
s. Promotes wholesome student-teacher relationships through fairness, understanding, and cheerfulness.					
t. Gives examinations sufficiently frequently for use in teaching as well as for grading purposes.					
u. Strives to be tactful in the use of constructive criticism.					
v. Gives proper consideration to the ideas of others.					
w. Admits his own mistakes.					

FIGURE 19B. Crowded and Poorly Reproduced Questionnaire.

Figure 20B. The responses for the items on the other side of the card, as seen in Figure 20B, can be indicated to the left of the black line on the business reply card. This questionnaire was sent out to school systems in bulk mailing and the individual

Please check these items in the appropriate spaces.

Sex.......................

Marital Status.......................

Dependents, other than self.........

Types of schools attended...........

Years training.......................

No. years experience.......................

Grade levels now teaching.............

Grade levels prepared to teach.....

Subject areas now teaching.............
(Secondary teachers only)

Subject areas prepared to teach...
(Secondary teachers only)

No. teachers in school district.......

DIRECTIONS: In column I, check the degree of competence you believe you possess of the following phases of your work. Then in Column II, check the types of training you believe would further your competency most. If you believe none would contribute, do not check any. Check several if you believe they contribute equally.

General methods of teaching.......................
Command of subject matter.......................
Directing pupil-participation in planning their work.........
Using audio-visual materials and equipment.................
Using newer-type reference materials.......................
Using newer techniques of teaching.......................
Participation in guidance activities.......................
Directing pupil evaluation.......................
Supervising extra-curricular activities.......................
Preparing lesson plans, units, courses, etc.................
Working with parents and laymen.......................
Working with colleagues and administrators.................
Contributing to general morale and supporting ethics.........

Oregon In-Service Professional Growth Study

Sponsors: The State Department of Education and the State System of Higher Education.

Purpose: To determine the present status of, and opinions regarding, in-service training and professional growth of Oregon teachers as a basis for future planning.

Directions: Do not sign your name. Check all answers. Tear off card and mail individually, or in groups in an envelope. Return card only, not questions.

COMMENTS:

Do not mail after Feb. 15, 1950

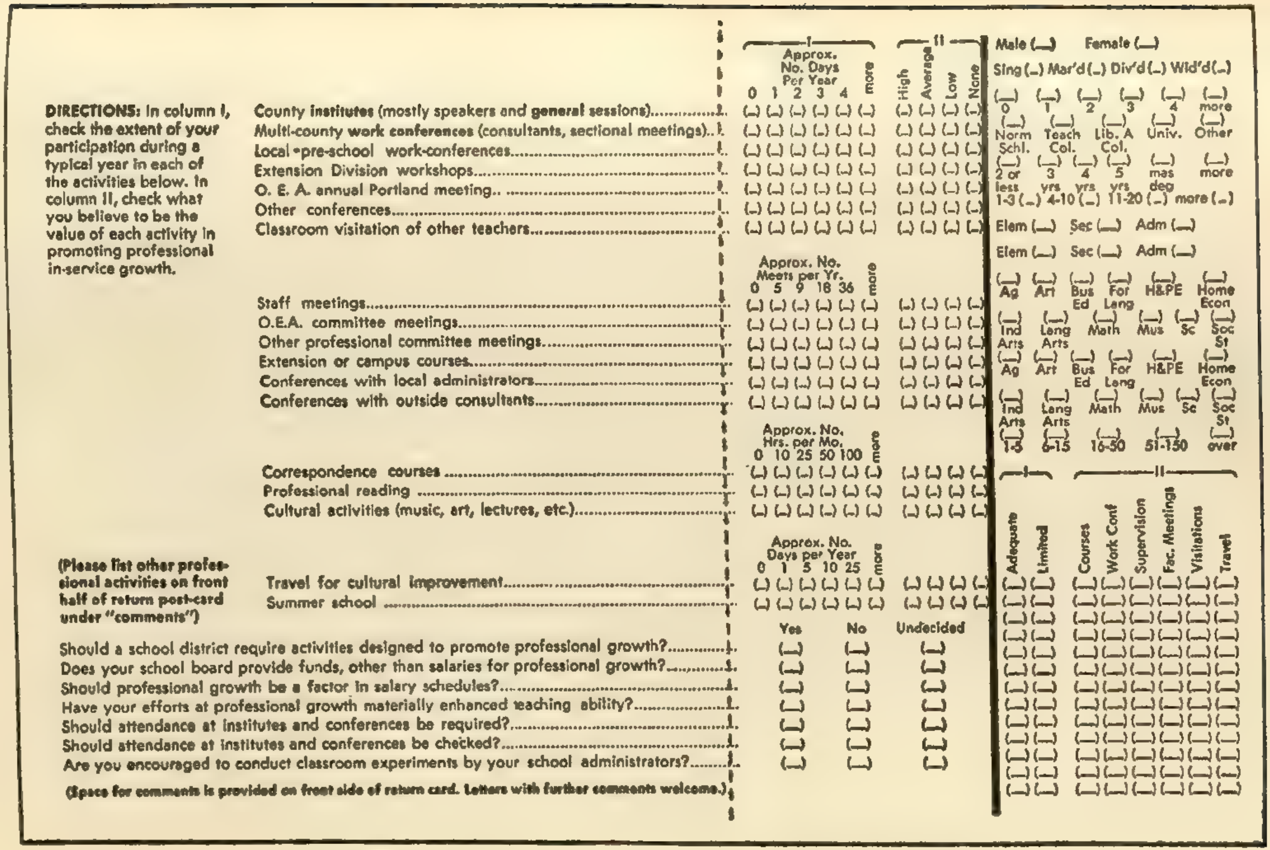
FIRST CLASS
PERMIT No. 71
(Sec. 34.9, P.L.&R.)
Eugene, Oregon

BUSINESS REPLY CARD
No Postage Stamp Necessary If Mailed in the United States

POSTAGE WILL BE PAID BY—

General Extension Division
Eugene, Oregon
Oregon State System of Higher Education

FIGURE 20A. A Carefully Planned Postal Card Questionnaire (front).



DIRECTIONS: In column I, check the extent of your participation during a typical year in each of the activities below. In column II, check what you believe to be the value of each activity in promoting professional in-service growth.

	I: Approx. No. Days Per Year: 0	1	2	3	4	more	II: High	Average	Low	None
County institutes (mostly speakers and general sessions)	( )	( )	( )	( )	( )	( )	( )	( )	( )	( )
Multi-county work conferences (consultants, sectional meetings)	( )	( )	( )	( )	( )	( )	( )	( )	( )	( )
Local pre-school work-conferences	( )	( )	( )	( )	( )	( )	( )	( )	( )	( )
Extension Division workshops	( )	( )	( )	( )	( )	( )	( )	( )	( )	( )
O. E. A. annual Portland meeting	( )	( )	( )	( )	( )	( )	( )	( )	( )	( )
Other conferences	( )	( )	( )	( )	( )	( )	( )	( )	( )	( )
Classroom visitation of other teachers	( )	( )	( )	( )	( )	( )	( )	( )	( )	( )
	Approx. No. Meets per Yr.: 0	5	9	18	36	more				
Staff meetings	( )	( )	( )	( )	( )	( )	( )	( )	( )	( )
O.E.A. committee meetings	( )	( )	( )	( )	( )	( )	( )	( )	( )	( )
Other professional committee meetings	( )	( )	( )	( )	( )	( )	( )	( )	( )	( )
Extension or campus courses	( )	( )	( )	( )	( )	( )	( )	( )	( )	( )
Conferences with local administrators	( )	( )	( )	( )	( )	( )	( )	( )	( )	( )
Conferences with outside consultants	( )	( )	( )	( )	( )	( )	( )	( )	( )	( )
	Approx. No. Hrs. per Mo.: 0	10	25	50	100	more				
Correspondence courses	( )	( )	( )	( )	( )	( )	( )	( )	( )	( )
Professional reading	( )	( )	( )	( )	( )	( )	( )	( )	( )	( )
Cultural activities (music, art, lectures, etc.)	( )	( )	( )	( )	( )	( )	( )	( )	( )	( )
	Approx. No. Days per Year: 0	1	5	10	25	more				
Travel for cultural improvement	( )	( )	( )	( )	( )	( )	( )	( )	( )	( )
Summer school	( )	( )	( )	( )	( )	( )	( )	( )	( )	( )

(Please list other professional activities on front half of return post-card under "comments")

	Yes	No	Undecided
Should a school district require activities designed to promote professional growth?	( )	( )	( )
Does your school board provide funds, other than salaries for professional growth?	( )	( )	( )
Should professional growth be a factor in salary schedules?	( )	( )	( )
Have your efforts at professional growth materially enhanced teaching ability?	( )	( )	( )
Should attendance at institutes and conferences be required?	( )	( )	( )
Should attendance at institutes and conferences be checked?	( )	( )	( )
Are you encouraged to conduct classroom experiments by your school administrators?	( )	( )	( )

(Space for comments is provided on front side of return card. Letters with further comments welcome.)

Male ( ) Female ( )

Sing ( ) Mar'd ( ) Div'd ( ) Wid'd ( )

( ) 0 ( ) 1 ( ) 2 ( ) 3 ( ) 4 ( ) more

( ) Norm Schl. ( ) Teach Col. ( ) Lib. A Col. ( ) Univ. ( ) Other

( ) 2 or less ( ) 3 yrs ( ) 4 yrs ( ) 5 yrs ( ) mas deg ( ) more

1-3 ( ) 4-10 ( ) 11-20 ( ) more ( )

Elem ( ) Sec ( ) Adm ( )

Elem ( ) Sec ( ) Adm ( )

( ) Ag ( ) Art ( ) Bus Ed ( ) For Lang ( ) H&PE ( ) Home Econ

( ) Ind Arts ( ) Lang Arts ( ) Math ( ) Mus ( ) Sc ( ) Soc St

( ) Ag ( ) Art ( ) Bus Ed ( ) For Lang ( ) H&PE ( ) Home Econ

( ) Ind Arts ( ) Lang Arts ( ) Math ( ) Mus ( ) Sc ( ) Soc St

( ) 1-5 ( ) 6-15 ( ) 16-50 ( ) 51-150 ( ) over

I: Adequate	Limited	II: Courses	Work Conf	Supervision	Fac. Meetings	Visitations	Travel
( )	( )	( )	( )	( )	( )	( )	( )
( )	( )	( )	( )	( )	( )	( )	( )
( )	( )	( )	( )	( )	( )	( )	( )
( )	( )	( )	( )	( )	( )	( )	( )
( )	( )	( )	( )	( )	( )	( )	( )
( )	( )	( )	( )	( )	( )	( )	( )
( )	( )	( )	( )	( )	( )	( )	( )
( )	( )	( )	( )	( )	( )	( )	( )
( )	( )	( )	( )	( )	( )	( )	( )
( )	( )	( )	( )	( )	( )	( )	( )
( )	( )	( )	( )	( )	( )	( )	( )
( )	( )	( )	( )	( )	( )	( )	( )
( )	( )	( )	( )	( )	( )	( )	( )

FIGURE 20B. A Carefully Planned Postal Card Questionnaire (back).

cards and covering letters were distributed to teachers by the school principals. In this manner the cost of sending out the survey materials was considerably less than it would have been if separate letters were sent to all the teachers individually. The business reply card made it necessary for the researcher to pay postage only for the information returned to him. This is the best designed postal card type of questionnaire that this writer has seen to this time.

If the questionnaire is typed, or if a process based on typing is used, an electric typewriter, some of which are equipped with boldface type or italics, provides a more effective typography for eliciting responses than a manual machine. Although most questionnaires are mimeographed or dittoed, nothing surpasses printing for choice of type forms and ease of arranging material impressively on a page. However, the researcher should investigate other methods of reproduction, such as multilith and various photo-offset processes. Some of these use paper, plastic, or metal "masters" which, when typed on electric typewriters, closely approximate the appearance of printing.

The cost of nearly all processes should be investigated and the amount of postage determined for each before making the final decision as to whether the mailed questionnaire is to be mimeographed or printed. Since a printed questionnaire takes only one-third to one-fourth as many pages as a mimeographed one, the cost of postage in many instances may offset the additional cost of printing, especially if the number of copies is great. It should be kept in mind that printing costs are largely in setting the type, and the added expense of running off a few hundred more questionnaires is negligible.

The paper used should be of high quality. If convenient, it should be the same size as that which will be used in the final research report so that copies can be included as part of the appendix. If the final report is to be typed as a thesis or dissertation, the paper should be 8½ x 11 inches in size, and the margins should be the same for the mailed questionnaire as for the context pages of the thesis.

Group Questionnaire Technique

Some of the difficulties encountered in obtaining good responses to questionnaires can be overcome by using the "group technique." This involves meeting with groups of respondents and having all members of each group fill in the questionnaire at the same time. This is especially well adapted to studies of teachers at regularly scheduled faculty meetings, club members, and the like. The researcher meets with the group, distributes the questionnaires, explains the purpose of his investigation, describes the type of data sought by the questionnaire, and answers questions about the study or specific items. In this manner he avoids some of the misunderstandings that oftentimes make responses to the mailed questionnaire questionable.

The major difficulty with this technique is that it limits the use of the questionnaire to those types of items that can be answered offhand or in a relatively short time and without much thought. It also assumes that all members of the group are able to answer the items accurately within the time limit allowed for this purpose, but such an assumption is not consistent with knowledge about differences among individuals. Sometimes the group may regard the researcher as an intruder if he has not arranged for official support for his investigation. Then, too, the time he takes having the group answer the questionnaire may be considered to infringe on other more desirable activities of the group or upon their leisure time.

On the other hand, this type of data collection is the least costly of all techniques because it eliminates the creation of a mailing list, addressing envelopes, postage, and tracing. It also provides an opportunity to clarify questions raised by one individual for the benefit of others who might be too self-conscious to ask or simply unaware that they did not understand an item. It gives the researcher an opportunity to arouse strong interest in the investigation and thus to improve the quality of responses.

PRETESTING THE QUESTIONNAIRE

In the preparation of a questionnaire, it is usually wise to formulate a preliminary draft and to secure expert criticism of its make-up and content, as many badly constructed questionnaires tend to annoy the recipients. The preliminary draft should be submitted to a few individuals similar to those who are to receive it eventually. Ask them to fill it in and, while it is still fresh in their minds, discuss it with you. In this way ambiguous questions, questions using unfamiliar words, or words that are likely to cause confusion or misunderstanding can be brought to your attention. Questions which the try-out group tend to omit or to answer superficially can then be revised, and the whole questionnaire may be improved. Alternate forms of phrasing difficult questions may also be tried out to determine which provides the most accurate responses.

In editing the try-out returns, items should be cross-checked for consistency of responses in different parts of the questionnaire. If there are responses that are contrary to known facts, attempt to find out why such responses were made. Revise questions for which more than one answer was given when only one answer should apply. Also, determine whether the categories for analyzing the data are satisfactory, or whether they need to be revised because there is too little data in some and too much in others.

PRELIMINARY QUESTIONNAIRE ANALYSIS

The interpretation of questionnaire returns and preparation for their analysis should be planned *in advance* of the time the questionnaires are distributed to the respondents. Several pertinent questions need to be considered on the basis of the preliminary try-out, and modifications should be made to avoid receiving returns that are not amenable to satisfactory analysis.

What are "usable" returns?
What should be done about nonrespondents?
What meanings can be attached to incomplete returns?

How can obvious inconsistencies be analyzed?
How can "no information" items be analyzed?
What kind of tabulations or summaries should be made?
What kinds of analyses should be made of these tabulations?

Questionnaires that are filled in completely, have no obvious inconsistencies, and are sufficient in number to provide an unbiased representation of the population being studied can certainly be considered as "usable" returns. If some of the information requested has been omitted by some of the respondents but not by others, there is question about the usability of the returns. For example, age may be given by some respondents but not by others, income by some but not by others, opinions about administrative practices by some and not by others. In such cases a mass of data may result, but they cannot be used for the analysis of the *entire* group. On the other hand, incomplete returns are not entirely useless; there may be a sufficient number of items for which information has been provided by *all* respondents to analyze those items for the entire group; and a sufficient number of items for which information has been provided by a *sizable number,* but not all, of the respondents to make analysis of those items for a subgroup.

It often happens that questionnaires are not returned. This may be because they have been sent to wrong addresses or to respondents away from home for the time being or who lack sufficient interest in the study to "bother" to respond. This "nonresponse" group constitutes an important practical problem. Unless it happens to be a small proportion of the whole sample, the results of the study are likely to be invalid. To obtain the delinquent information, however, may require much time and money, and it may be impossible to obtain in many cases; but every effort should be made to reduce the nonresponses to negligible proportions by using the techniques discussed in the previous section of this chapter. To ignore the nonresponse group may result in a sample that has a *bias of unknown magnitude.*

If the required information cannot be secured from the non-

PRETESTING THE QUESTIONNAIRE

In the preparation of a questionnaire, it is usually wise to formulate a preliminary draft and to secure expert criticism of its make-up and content, as many badly constructed questionnaires tend to annoy the recipients. The preliminary draft should be submitted to a few individuals similar to those who are to receive it eventually. Ask them to fill it in and, while it is still fresh in their minds, discuss it with you. In this way ambiguous questions, questions using unfamiliar words, or words that are likely to cause confusion or misunderstanding can be brought to your attention. Questions which the try-out group tend to omit or to answer superficially can then be revised, and the whole questionnaire may be improved. Alternate forms of phrasing difficult questions may also be tried out to determine which provides the most accurate responses.

In editing the try-out returns, items should be cross-checked for consistency of responses in different parts of the questionnaire. If there are responses that are contrary to known facts, attempt to find out why such responses were made. Revise questions for which more than one answer was given when only one answer should apply. Also, determine whether the categories for analyzing the data are satisfactory, or whether they need to be revised because there is too little data in some and too much in others.

PRELIMINARY QUESTIONNAIRE ANALYSIS

The interpretation of questionnaire returns and preparation for their analysis should be planned *in advance* of the time the questionnaires are distributed to the respondents. Several pertinent questions need to be considered on the basis of the preliminary try-out, and modifications should be made to avoid receiving returns that are not amenable to satisfactory analysis.

What are "usable" returns?
What should be done about nonrespondents?
What meanings can be attached to incomplete returns?

How can obvious inconsistencies be analyzed?
How can "no information" items be analyzed?
What kind of tabulations or summaries should be made?
What kinds of analyses should be made of these tabulations?

Questionnaires that are filled in completely, have no obvious inconsistencies, and are sufficient in number to provide an unbiased representation of the population being studied can certainly be considered as "usable" returns. If some of the information requested has been omitted by some of the respondents but not by others, there is question about the usability of the returns. For example, age may be given by some respondents but not by others, income by some but not by others, opinions about administrative practices by some and not by others. In such cases a mass of data may result, but they cannot be used for the analysis of the *entire* group. On the other hand, incomplete returns are not entirely useless; there may be a sufficient number of items for which information has been provided by *all* respondents to analyze those items for the entire group; and a sufficient number of items for which information has been provided by a *sizable number,* but not all, of the respondents to make analysis of those items for a subgroup.

It often happens that questionnaires are not returned. This may be because they have been sent to wrong addresses or to respondents away from home for the time being or who lack sufficient interest in the study to "bother" to respond. This "nonresponse" group constitutes an important practical problem. Unless it happens to be a small proportion of the whole sample, the results of the study are likely to be invalid. To obtain the delinquent information, however, may require much time and money, and it may be impossible to obtain in many cases; but every effort should be made to reduce the nonresponses to negligible proportions by using the techniques discussed in the previous section of this chapter. To ignore the nonresponse group may result in a sample that has a *bias of unknown magnitude.*

If the required information cannot be secured from the non-

respondents of the population sampled, there is no amount of statistical ingenuity capable of providing the researcher with information that may be considered representative of the population originally sampled. However, by being able to identify each questionnaire returned (see pp. 141–146), the researcher will be able to describe accurately the sample obtained and to limit his conclusions to the type of population represented by such a sample.

Incomplete returns can be analyzed only on the basis of those items for which data are provided. It is not possible for a researcher to use any responses that appear to be contradictory or inconsistent without a follow-up to determine exactly what the respondent had in mind. As soon as the first questionnaires start coming in, it is desirable to check them to identify those which need to be followed up because of apparent errors, inconsistencies, and incomplete returns before the respondent has forgotten what he intended to say or before he has become unavailable for a follow-up.

It is also necessary to determine how to interpret and differentiate between responses that are marked on a questionnaire with a "zero" and those that are omitted. In analyzing the data, the "no information" or omitted items cannot be averaged in with the others in the group. On the other hand, when "zero" is used to indicate a value, rather than no information, it cannot be excluded in averaging the values or in other techniques of summarization. In tabulating the data, "frequency counting" of items is not always sufficient. It is often necessary and desirable, in addition, to make a full interpretation of the returns which involves making an evaluation of the questionnaire as a whole and taking into consideration possible interrelationships among the various items.

Skeletal tables should be prepared, prior to sending out the questionnaires, to see how the data requested can be presented in order best to answer the questions under investigation. In addition, the types of statistical analyses that are to be used in the interpretation of the data should be determined at this time. Some researchers have found it profitable to fill in the

skeletal tables with fictional data like that expected from the questionnaires to be returned and to try out the statistical procedures for analyzing such data. If the results of these tabulations and analyses seem to provide for the kinds of conclusions that could be meaningfully interpreted, the researcher has some assurance that his plans for analysis are adequate. Then, when the actual data start coming in the researcher knows exactly what he is to do with them and how to make his analyses.

In this discussion of questionnaires, the topic of what the researcher should do if he does not receive replies from his complete sample has been omitted purposely. This is a serious and controversial problem in statistical theory which goes beyond the scope of this text. Adequate discussions can be found in most texts devoted to sampling techniques.

SELECTED REFERENCES

Ahrens, Edward H. J. "Validity of the Questionnaire," *Science Education, 34:*41–42, February, 1950. (Presents criteria for evaluating a questionnaire.)

Deming, W. Edwards. "On a Probability Mechanism to Attain an Economic Balance Between the Resultant Error of Response and the Bias of Non-response," *American Statistical Association Journal, 48:*743–72, 1953.

El-Badry, M. A. "A Sampling Procedure for Mailed Questionnaires," *American Statistical Association Journal, 51:*209–27, 1956.

Fischer, R. P. "Signed Versus Unsigned Personal Questionnaires," *Journal of Applied Psychology, 30:*220–225, 1946.

Goode, William J., and Hatt, Paul K. *Methods in Social Research*. New York: McGraw-Hill, 1952. 386 pp. (Read chap. 11.)

Hansen, Morris H., and Hurwitz, William N. "The Problem of Non-response in Sample Surveys," *American Statistical Association Journal, 41:*517–29, 1946.

Huffman, Harry A. "Improving the Questionnaire as a Tool of Research," *National Business Education Quarterly, 17:*15–18, October, 1948. (Presents procedures to help persons construct scientifically useful questionnaires.)

Mark, Mary Louise, *Statistics in the Making*. Columbus, Ohio: Bu-

reau of Business Research. Ohio State University, 1958. 436 pp. (See especially pp. 231–257.)

Politz, Alfred, and Simmons, Willard R. "An Attempt to Get the Non-at-homes into the Sample Without Call-backs," *American Statistical Association Journal, 44*:9–31, 1949.

CHAPTER VII

Documentary Analyses and Other Techniques of Data Collection

Documentary Analyses
- General sources
- Theses
- Government documents
- Library services
- Types of documentary analyses
- Categorization of information
- Critical analysis of data sources and documents
- Critical analysis of the documents

Combinations of Methods of Data Collection
- Observation interview method
- Patterned interview
- Group interview-questionnaire method
- Case-study method
- "Action research"

In addition to using the techniques for collecting information directly from the persons involved, as described in the preceding three chapters, the researcher also makes extensive use of material in published form. These materials may include various statistical data collected by governmental agencies, private institutions, or individuals, and also many other types of docu-

mentary information. This chapter describes some of the sources of materials typically used in educational research and the principles and techniques for documentary analysis.

DOCUMENTARY ANALYSES

Documentary analyses in research deal with records that already exist, and such analyses have often been referred to as a method of library research, historical research, or "content analysis." These analyses are primarily quantitative in nature, dealing with characteristics that can be identified and counted, although qualitative interpretations are often necessary and desirable to determine what to count and to be able to define them. Their major importance in educational research is in curriculum development in which a researcher may make a survey of what is going on, or what has taken place, as reported in written or printed materials. The following three broad categories of literary sources of data are frequently used for collecting data.

1. *Factual sources,* including institutional catalogues; records of accounts, meetings, and events; laws and official documents; and news articles.
2. *Educational implements,* including courses of study and reading lists, syllabi, teachers' manuals, and pupils' textbooks and workbooks.
3. Sources dealing with the *interpretation and study of education,* including reports of research studies, reports of the point of view of individuals or committees, and textbooks or other treatises on the subject of education.

In Chapter II, several published references were identified as sources for the development of a research project. These may also be used as aids in the search for materials for documentary analyses. The researcher will also find that the Library of Congress, the New York Public Library, and specialized libraries in some universities or municipalities have valuable collections in particular educational areas. Some state and local historical societies have museums that contain important edu-

cational remains, newspaper files, and documents. And many private citizens and prominent educators have accumulated exceptional collections of valuable documentary data. However, many source materials are not conveniently collected for a researcher. He may have to conduct several probing expeditions to find usable evidence, such as talking with "old-timers"; browsing in book shops, attics, and second-hand stores; checking through the files of minutes and correspondence of school systems, teachers, and professors; and studying institutional reports and records.

Since the great bulk of library information required by most researchers is in text form, this section provides suggestions on how to compile a bibliography in addition to those made in Chapter II. Library research is involved in all research, even though the data used are empirical rather than literary so that this section is applicable to all research studies.

An author's reasoning, conclusions, assumptions, and data may be evaluated. Many writers in the same field may be compared or contrasted with each other for the purpose of deriving an eclectic explanation. Or, existing writings may be reexamined in the light of information made available since their publication. Although this sort of research seldom uses mathematical tools for analysis, nevertheless the analysis should be rigorous and exacting.

General Sources

It is assumed that the reader is familiar with the card catalog in the library in which he does his work. He should pay attention to the subject headings on the cards, particularly those ending with the word "bibliographies." There is no use in compiling a bibliography from scratch, if much of it can be taken from one already compiled. Each book which may be relevant should be entered on a card; then the investigator should carefully scan the book and record all information that may be of value to him. He should watch for footnote references in these books to other books and articles on the same subject which may be added to the bibliography he is building. He

should be especially careful to make his bibliographic references in complete form; nothing is so frustrating as to find that one needs to cite a reference in a footnote, but that one lacks the complete information.

Theses

Frequently a good source of bibliographical information is the unpublished theses which can be obtained only through the library at the college or university at which the writer received his degree. The researcher should have no difficulty in locating appropriate theses in his own university library, or in near-by libraries. However, if the theses are located in libraries which the researcher cannot visit conveniently, he may easily be unaware of their existence. This means that he is dependent upon some form of listing to locate theses in libraries at a distance.

Bachelors and masters theses may be pretty well written off because they are seldom listed anywhere except in the institution at which the degree was granted. This does not mean that they would not be of value if the researcher could locate them; they are practically impossible to find.

Doctoral theses can be located in a number of ways. Fairly recent theses are listed in *Dissertation Abstracts,* published monthly by University Microfilms, Inc., Ann Arbor, Michigan. These monthly releases are accumulated into annual volumes. This service lists doctoral dissertations written at more than 100 graduate schools in the United States. It is possible to obtain complete copies of the theses on microfilm or on pages 5½ x 8½ inches large which are processed by the Xerox process. However, the coverage by institution is poor because the particular school or department in which the thesis is being written must forward the information to *Dissertation Abstracts* in order for it to be included.

Government Documents

A vast amount of material is constantly coming out of federal departments, agencies, and Congress. These government documents may prove to be invaluable to a researcher. Unfortun-

ately, locating the proper publication is frequently a very difficult task. In the first place, only the larger libraries are government depositories. Even these libraries find that they must constantly check to see that they actually do receive all the publications to which they are entitled. Assuming that the researcher has access to a library that has federal government documents, the next problem is locating them. Discovering the title in the card index is quite a task in itself. In fact, the search for the proper card, in order to obtain the call number, is an assignment which frequently requires the aid of the government documents librarian. The second difficulty in locating government documents is learning of their existence. How the researcher finds out about government publications which may be of help to him is the subject of the remainder of this section.

Aside from laws and court opinions, which will not be discussed in this book despite their importance to researchers, government publications are of the following types:

1. Hearings before Congressional committees
2. U.S. House or Senate Documents
3. Congressional Committee prints
4. U.S. House or Senate reports
5. Publications of U.S. executive or independent agencies

Hearings are meetings of a committee of the House of Representatives, the Senate, or a joint committee to hear testimony on a given subject. These may be for all sorts of purposes such as the following example of hearings (in proper bibliographical form).

U.S. Congress. Senate. Special Committee on Unemployment Problems. Unemployment Problems, Part 8, hearings, 86th Congress, November–December 1959. Washington, U.S. Gov't. Print. Off., 1960.

Documents are prepared by outside agencies and reprinted by a committee or by one of the houses of Congress, as indicated in the following example.

U.S. Congress. House. Foreign Trade Interests of the State of Oregon. House Document No. 232, prepared at the request of the

Oregon Congressional delegation by the Legislative Reference Service of the Library of Congress, the Oregon State System of Higher Education, and the Oregon State Department of Planning and Development, October 1959. Washington, U.S. Gov't. Print. Off., 1959.

Prints are similar to *documents*, except that they are prepared by staff members of a Congressional agency. House and Senate *Reports* are self-explanatory, as are agency publications.

The first step in finding out about a federal government publication is to refer to the *Catalog of United States Government Publications* which is issued monthly and accumulated into annual volumes. The catalog is classified according to the department or bureau responsible for the publication. This classification will not help the student unless his subject matter is logically in the jurisdiction of one particular agency. Probably the most valuable aspect of the Catalog is the index which is arranged by subject. The researcher should be warned, however, that the subject approach in the index can also be treacherous because a hearing in which one witness discusses certain policies of his agency—a subject which may be almost impossible to investigate directly because of the officers' reluctance to talk with the researcher—will be indexed under the principal topic of the hearing, and therefore the researcher has no way of knowing that there is testimony relating to the policies in which he is interested.

The growth during the last few years of government publications of all kinds, especially House and Senate committee *documents*, has been little short of phenomenal. There are very few research problems for which relevant material cannot be found in some government publication.

Library Services

Large libraries offer a number of services to researchers. Oftentimes graduate students flounder aimlessly because they are not aware of the help that is available to them. Many libraries have some or all of the following services:

1. An interlibrary loan service which permits the student to bor-

row books, theses, and documents from other libraries. This may be most useful, although students must not expect to depend upon other libraries for more than an occasional source.

2. "Vertical Files," as the librarians call them. These are pamphlets and leaflets that are kept in office-type files and are not entered in the card catalog because they are considered to be of ephemeral interest only. The appropriate librarian can help the student find this material.
3. Subject specialists on their staff, a part of whose job is to help students learn how to find items. A government document specialist is almost a necessity if the library is a government depository. A Social Science specialist can be of great aid to a student in educational research.
4. A system for calling in books that are in circulation. However, the student should exploit the book thoroughly while he has it because he cannot expect to keep it indefinitely as other researchers may want it.

Types of Documentary Analyses

A wide variety of documentary studies are made in educational research; such as analyses of textbooks, story-content studies, and the examination of periodical or newspaper content. *Textbook analyses* include such determinations as the frequency of concepts, the kinds of concepts, classification of techniques, tracing of historical trends, status studies based on definite criteria, and studies of errors, distortions, and biases in printed materials. *Story-content* studies may involve, for example, the identification of ethnic backgrounds and the roles which characters play in literature. Such studies might be made to determine if there is a demonstrable discrimination against certain types of people. Studies of *newspaper content* might involve the counting of the frequency of appearance of certain symbols such as "United Nations," "communism," "totalitarianism," "capital," "labor," "democracy," "Jews," "Russia," "England," and the like. It might determine whether the presentation of ideas is favorable, neutral, or unfavorable. Such studies serve as bases for the investigation of many social problems.

Documentary analysis is the principal method used in "historical research," which may be defined as an attempt to find out what has happened in the course of time and to correlate the events, within the limits of available materials on the one hand and of the researcher's intelligence and understanding on the other, into a meaningful sequence. The first task of the historian, according to Kohn, is to

> . . . find out by patient and painstaking research the true facts of the past. . . . history is scholarly research. . . . Its responsibility is to find out how it was, not how it should have been. Yet no historian can know the whole past, not even the full story of one man, or one year, or even of one day. From the infinity of facts we are always forced to select within the limits of surviving documentation and those imposed by our intelligence and our intentions. We select according to our set of principles. . . . Historians can teach us not to look on nations and classes as isolated phenomena but to see and judge them in a universal context in the light of this ethical tradition. . . . each generation rewrites history not by adjusting facts to an alleged need of the hour, but by changing its viewpoints, as an old man looking back on some period of his youth, though he has achieved greater factual knowledge, arrives at a different judgment as a result of his experience.[1]

Categorization of Information

Documentary analysis requires that categories of data be clearly and explicitly defined so that other researchers can apply them to the same content in order to verify conclusions. In this respect, some categories may have to be set up on *a priori* bases, or by logical considerations, in advance of inspecting the documents. This procedure gives direction to the researcher in examining documents. On the other hand, some categories may be set up on *a posteriori* bases, or categories derived from the specific materials examined. While the latter procedure requires a considerable amount of trial and error experimentation with different possible schemes, and depends on the ability of the researcher to single out significant bits of information and to work out a practical method of classification, it often pro-

[1] Hans Kohn, "A Historian's Creed for Our Time," *American Association of University Professors Bulletin* (Winter, 1953–1954), *39*:608–615.

vides for better categories than the *a priori* method because of inadvertent omissions or impractical classifications that may have been made by the researcher. In using content analysis in the solution of a topic under investigation, the researcher must classify methodically all materials that are relevant to the topic and avoid omitting any that are not interesting to him. He may also use quantitative procedures to permit comparisons to be made with other materials.

Critical Analysis of Data Sources and Documents

When using written and printed materials, it is necessary to find out the conditions which may have affected their composition and content. In some instances, these may be so well known that there is no necessity for making a careful individual review. Nonetheless, the researcher should always remember that documents may not be the accurate source which they purport to be.

The proceedings of a conference are often particularly treacherous documents. The discussion following the talks, and often even the talks themselves, are not written before delivery and are transcribed from a tape and then heavily edited, frequently by a secretary who is relatively uninformed on the subject matter covered. Such a process allows errors to be introduced at three points: (1) the speaker may be careless in his statements because he is speaking conversationally, (2) the person transcribing the tape frequently makes mistakes, and (3) the person editing the transcript may alter the meaning of the speaker. Approval of the text by the speaker before publication will reduce the chance for error, but this is a procedure infrequently followed. Inaccuracies in such documents are unfortunate because comments made at meetings may be almost impossible to duplicate.

External criticism. This type of criticism deals with the *authenticity* or the genuineness of the document—whether it is what it appears or claims to be—to determine whether it is admissible as evidence. It is concerned with the form and appearance of the document rather than meaning of the contents, although external criticism at times may employ internal evidence from the document through a study of its contents.

Internal criticism. This deals with the *credibility* or meaning and trustworthiness of the data within the document. It is concerned with establishing the time, place, and authorship of the document and restoring the original form and language employed by the author. Many words in older documents do not mean the same thing today that they did in earlier times.

There is no sharp line of demarcation between internal and external phases of criticism, and the two processes may progress simultaneously with considerable overlapping. The following is an excellent list of questions to ask when analyzing the sources of a document:

1. *Where was it produced?* Usually, with printed materials, the publisher's location is given. But it is more important to know where it was written and under what conditions than just to know who published it and where it was published.

2. *When was it written?* What time gap was there between the time it was written and the date of publication? If no date is given, can it be determined from the contents of the document?

3. *Is it a valid document?* Is the document original or is it a copy? Is it a sound or corrupt text due to either error or fraud? Is it an absolutely accurate copy or has it been "edited"? Is it in complete form or has it been partially destroyed?

4. *Who was the author?* What is known about the author's nationality, vocation, position, class, party affiliations, religion, training, and so forth? What qualifications such as training, mental characteristics, social status, biases, interests, and linguistic habits did he have for writing the document? Is the authorship true or false? Was the document written by a "ghost writer"? Does the document represent a case of plagiarism or forgery?

5. *May the document be accepted as true?* The following questions should be considered in determining the meaning, honesty, and accuracy of the author. What is the real meaning? Why did the author write the document? Does the writer show a distortion of facts resulting from vanity? Was the author placed in a position which would invite him to deviate from the truth? Does the desire for promotion cause him to write to please his superiors? Does the author labor under exaggerated political bias? Is he expressing sentiments to please the public? Has he violated the truth through the

use of literary artifices? Was he a good observer? When did the author record his observations? Are the facts of such a nature that they could not be learned by observation alone? Was the author being questioned? To what extent may anonymous statements, or author unknown, be interpreted? Are some facts so well known that it would be difficult to make any errors concerning them? Are the facts of such a nature as to render falsehood improbable?[2]

Critical Analysis of the Documents

If the researcher is satisfied, after critical analysis, with the sources of the documents he is studying, his next step is to seek out the facts and conceptions that are pertinent to his topic for research. He must then determine what relationships exist among the various facts, make comparisons with facts presented by various authors, and attempts to establish truths.

When independent statements of facts disagree, it is necessary to resolve the inconsistencies. In this respect a researcher should use other means than that of accepting only the views or statements of a majority of writers.

> Where a majority of observers take one point of view and a minority another, it does not follow that the critic should accept the majority view. If one group is right, it follows that the other group is wrong, and it may well be that the majority is wrong. . . . The opinions of two persons who are well acquainted with a fact are worth more than the opinions of scores who know little or nothing of it. Of course, it is by external criticism that one group should be eliminated; but if the critic has been unable to do this he cannot draw conclusions, he can only state opinions as they exist and cite his authorities.[3]

COMBINATIONS OF METHODS OF DATA COLLECTION

The preceding four chapters have discussed the various methods of collecting information—observation, interview,

[2] Walter E. Spahr and Rinehart J. Swenson, *Methods and Status of Scientific Research*. New York: Harper & Row, 1930, pp. 92–127.

[3] *Ibid.*, pp. 131–132.

correspondence, and library or documentary. It is seldom that a research project will involve only one of these. Obviously, library research will be a part of almost every project. Combinations of other methods are almost always used to obtain data not already available in written form. The following are some of the combinations commonly used.

Observation Interview Method

Certain studies demand comparisons between what people actually do under given conditions and what they say or think they do. In such situations it may be convenient to collect data through interviews and to observe the persons interviewed in order to get at the "true" facts. This is one use of "participation observation" mentioned earlier.

Patterned Interview

The patterned interview combines the interview and questionnaire methods. A detailed schedule is prepared in order to insure coverage of each and every item for which information is needed. This tends to minimize differences in responses due to various methods of interviewing and is especially desirable when more than one individual is conducting the interviews. The amount of uniformity can be controlled by the degree of "open-endedness" in the schedule. The schedule may be very detailed, giving the respondent the choice of relatively few answers—a procedure which facilitates tabulation. On the other hand, the schedule may be so broad that it only provides the interviewer with the general topics to be covered in all interviews and permits the respondents to give very different answers—a procedure which makes tabulation difficult. Obviously, the latter is simply a structured interview whereas the former is a questionnaire read to the respondent. Most patterned interviews lie somewhere between these two extremes.

Group Interview-Questionnaire Method

The questionnaire method frequently results in a lack of

communication of ideas between the researcher and the respondent. This may be due to the poor phrasing of the questions or the inability of the respondent to understand the questions, or both. When interviewing, it is frequently possible for the questioner to detect when ideas are not being communicated and to reformulate his questions so that they are understood in the way he intended them to be. The group interview-questionnaire method involves the researcher meeting several individuals as a group, discussing with them the problem under investigation and the procedure he is using to attack the problem, and then asking each member of the group to answer the questionnaire. If the questions prove to be ambiguous, the researcher can clarify them exactly as in the interview situation. This procedure is best adapted to personnel and consumer research because they involve the sorts of situations that enable the researcher conveniently to gather a number of respondents in a group.

Case-Study Method

Students who have enrolled in even a few education courses have almost certainly become acquainted with case studies. A case study may deal not only with an individual but with almost any unit—a classroom, a school system, a city, or even an epoch in history. Its distinguishing characteristic is that it emphasizes the total situation. In so doing, it describes a situation or a sequence of events leading up to some particular behavior. The case-study method is to be contrasted with statistical methods which yield averages and in which departures from average in one direction tend to be offset by departures in the opposite direction.

Case studies imply an intensive investigation of a particular situation. One advantage of such an investigation is that it may expose a new relationship whose extent may then be measured by statistical techniques. Another advantage of such a study is that the researcher, if he selects a case which he believes to be typical of a class, may use his findings from this one case to form the bases for generalized hypotheses.

"Action Research"

Since there has been considerable publicity given to "action research," "educational engineering," and "operational research" during the past few years, it is being mentioned here although it is not strictly a method of collecting information and data. It is, rather, a program aimed at changing existing conditions step by step through group participation, primarily by teams of classroom teachers who find that the scientific method pure research workers employ to unravel the riddles of the universe can also aid them in solving immediate, local school problems. In this cooperative type of research, teachers describe current difficulties they experience in their classrooms (step 1). They ask trained research consultants to help them analyze these problems and carry out their investigations. Together, teachers and research consultants strive to bring the problems into focus by precisely determining the pertinent causal factors (step 2). From their study of the data, they construct hypotheses that seem to offer possible explanations for the causes of their difficulties (step 3). Subsequently they reason out the consequences of their hypotheses (step 4) and test the most appropriate hypothesis by an acceptable technique (step 5).[4] Several books and articles in professional journals have devoted considerable space to the discussion of "Action Research" and "Operations Research." One of the more comprehensive treatments of this subject has been published by Stephen Corey.[5]

[4] Deobold B. Van Dalen and William J. Meyer, *Understanding Educational Research*. McGraw-Hill, 1962, p. 25.

[5] Stephen M. Corey, *Action Research to Improve School Practice*. New York: Teachers College, Columbia University, 1953, 161 pp.

SELECTED REFERENCES

Anderson, Ruth I. "Action Research and the Classroom Teacher," *Business Education Forum, 14*:27–28, March, 1961.

Berelson, Bernard. *Content Analysis in Communication Research*. Glencoe, Ill.: Free Press, 1952. 220 pp. (Refer to index for specific readings.)

Dale, Edgar. "Quantitative Analysis of Documentary Materials."

Review of Educational Research, (December, 1939), *9*:466–471, 595–596.

Good, Carter V. *Introduction to Educational Research.* New York: Appleton-Century-Crofts, 1959. (See chap. 4.)

Goode, William J., and Hatt, Paul K. *Methods in Social Research.* New York: McGraw-Hill, 1952. 386 pp. (See chap. 9.)

McCormick, Thomas C., and Francis, Roy G. *Methods of Research in the Behavioral Sciences.* New York: Harper & Row, 1958. 244 pp. (See especially pp. 45–75.)

Van Dalen, Deobold, and Meyer, William J. *Understanding Educational Research.* New York: McGraw-Hill, 1962. (See pp. 163–173.)

CHAPTER VIII

Design of Research Projects

Illustrative Experimental Problems

Factors to Consider

Basic Types of Design Errors
- Type S (Sampling) errors
- Type G (Group) errors
- Type R (Replication) errors

Simple Randomized Design

Random-Replications Design

Groups-Within-Treatments Design

Other Designs

Treatment of Extraneous Variables
- Equalization
- Counterbalancing
- Randomization

The researcher must decide quite early in his work exactly how his project is to be carried out. This operating plan is called the project design. It should be carefully thought out so that a minimum of unpleasant surprises occur as it is being followed through. The project requires detailed design whether it is largely descriptive, utilizes library material entirely, or is

empirical and yields data which can be treated with advanced mathematical techniques. Obviously, the project design depends upon the particular technique of analysis to be used; and this depends in turn not only on the nature of the data, but also upon the abilities of the researcher.

Most empirical studies do permit some form of quantitative analysis. This chapter presents a very general description of several analytical approaches which affect the project design and some of the different ways in which designs can be developed. These are presented primarily to acquaint a beginning researcher with the *ideas* of the designs, but they are not discussed extensively enough to serve as models for actual operations. The actual use of the various designs discussed requires a more extensive study of statistical analysis than can be covered within the scope of this text.

In setting up plans for experimentation and statistical analysis it is desirable for the researcher to consider various designs to determine which would be appropriate for the experiments he may have occasion to perform and how to interpret the results obtained through the use of those designs. The major purpose of experiments is to describe the effect of certain treatments upon some characteristic of a group or population and to test some hypothesis about this effect. In planning experiments it is not only necessary to consider how to describe or measure the desired effect, but it is also necessary to consider how to test its significance. Frequently the experimental design used by many graduate students does not permit the use of any known test of significance, whereas a change could easily have been made in the original design to make this possible. A thoroughgoing treatment of this topic is beyond the scope of this book, but helpful discussions of experiment design and analysis are readily available elsewhere.[1]

The purpose of this section is to develop an awareness in the beginning researcher of the basic types of errors and the experimental designs that are used to take these errors into account

[1] Everet F. Lindquist, *Design and Analysis of Experiments in Psychology and Education*. Boston: Houghton Mifflin, 1953, pp. xix–393.

Only three basic designs are illustrated here: simple-randomized design, random-replications design, and the groups-within-treatments design. Many other more complex designs are readily available in the literature on experimental design, all of which may be considered as variations or combinations of these basic designs.

ILLUSTRATIVE EXPERIMENTAL PROBLEMS

For purposes of illustrating the basic types of errors and three experimental designs, the following two experimental problems are described in parallel form. One might well be a problem in poultry husbandry and the other a simple, restricted problem in education.

Poultry Husbandry	Education
Which of two different feeding procedures will result in the greater gain in weight of chickens? The problem is to determine the most efficient diet for increasing the weight of chickens from incubation to ten weeks of age.	Which of two different instructional methods will result in the greater gain in the skills of computation and understanding of the concepts in arithmetic of fifth grade pupils? The problem is to determine the most efficient method of developing arithmetic skills and concepts during the year pupils are in the fifth grade of school.

In carrying out either of these experiments, a *single-variable* design may be used. That is, the diet or method is the only variable that is being considered as the characteristic to be described and tested for a significant effect. The hypothesis that will be tested is that both diets or methods are equally good. If this hypothesis can be rejected at the 5 percent level of confidence, an inference can be drawn with respect to which diet or method is the more efficient of the two considered. To illustrate the use of three basic designs with the minimum of computational effort, the problem is restricted to a very small sample of cases. In an actual experiment, of course, the sample

would have to be much larger to avoid undue errors resulting from the sampling process. The single variable is expressed as treatments or methods A and B, and is measured in terms of gains in the means from the initial measures to the final measures.

Procedure: Select at random 60 newly hatched chicks from a large supply of chicks. Place 30 in one cage to be fed one diet, and 30 in another cage to be fed a different diet. Flip a coin to determine which cage of chicks is to be fed diet A and which is to be fed diet B. Obtain the mean weight of the chicks at incubation for each cage. At the end of the ten-week feeding experiment obtain the final mean weight and compute the mean gain in weight for each cage or diet. These will be indicated as M_a and M_b.

The hypothesis to test is: $M_a - M_b = 0$.

Procedure: Select at random 60 pupils from the fifth grade of a school. Assign 30 pupils to one class to be taught arithmetic by one method, and 30 pupils to a different class to be taught by another method. Flip a coin to determine which class is to be taught by method A, and which is to be taught by method B. Administer a criterion test to determine the scores of each pupil at the beginning of the experiment. Administer the test again at the end of the experimental time (one semester, for example). Compute the mean gain in scores for each class. These will be indicated as M_a and M_b.

The hypothesis to test is: $M_a - M_b = 0$.

FACTORS TO CONSIDER

In planning a single-variable experiment one should consider the many factors that might be involved and need to be controlled so that the effects of the treatments or methods can be tested for significance. In these illustrations, these factors should be controlled, if possible.

Poultry Husbandry

1. *Strain of chicks.* The chicks should all be drawn at ran-

Education

1. *Cultural background of pupils.* Pupils should be

POULTRY HUSBANDRY

dom from a specific strain of chicks to avoid any systematic difference in means due to one mean being unduly influenced by a predominance of chicks of a light-weight strain (e.g., Leghorn) and another of a heavy-weight strain (e.g., Plymouth Rock). This factor may be controlled by the experimenter.

2. *Sex.* Roosters might possibly gain more weight than hens from *any* type of diet. Therefore, the ratio of sex should be equalized for each diet group.
3. *Initial weight differences.* The sampling of chicks should be controlled to provide matched groups on initial weight, as this factor might influence the major effect of the two diets.
4. *Positions of cages in the experimental setting.* The cages should be located to equalize the effects of light, heat, drafts, commotion, and other factors on the two diets.
5. *Cleaning procedures in cages.* The time and manner of cleaning the cages should be equalized.

EDUCATION

drawn at random from those having approximately the same cultural backgrounds. If one class is composed primarily of children from one culture and the other of children from a different culture, the difference in means might be attributable to this factor rather than to differences in the methods of instruction.

2. *Sex.* Sex differences in the learning of arithmetic should be controlled by including an equal proportion of boys and girls in the two methods classes.
3. *Arithmetical aptitude.* Pupils should be matched in aptitude so that one class is not composed primarily of good students and the other of poor students.
4. *Location of classrooms.* The classrooms should be selected so that the effects of heat, light, and outside distractions are equalized.
5. *Avoidance of interruption.* Arrangements should be made so that classes are not disturbed by visitors, fire drills, and other types of distractions. However, if interruptions cannot be avoided, they should be equalized as much as possible

Poultry Husbandry

6. *Handling of diets.* The preparation of diets should be carefully controlled to avoid any errors which might unduly affect the experimental results, such as measuring out the daily portions of food, careless feeding schedule, poor care of cages, and others.

7. *Unapparent factors.* While the previous factors may be controlled, there are usually some factors that are unapparent at the beginning or during the experimental time. One cage could contain some chicks that had a low-order infection or undetected illness that would affect systematically the gain in weight of the chicks in one cage. One cage might also be influenced systematically by the presence of a "neurotic" chick. Factors which affect individual chicks, but not all the chicks in a cage, could be expected in other experiments and in the use of commercial diets so that their effects would be considered as part of the overall effects of the diets.

Education

6. *Control of methods.* The methods of instruction should be precisely determined for each class and followed explicitly. Classes should meet at the same time. Systematic differences may be present if one class meets early in the day and the other late in the day.

7. *Unapparent factors.* Experimental controls cannot be made for some factors such as the unapparent illness in a class in which several pupils have slight colds, or other low-order infections that might have an effect on their learning. The presence of one or more individuals who condition the attitudes of the pupils in the class as a whole, or of a "neurotic" child, may have a systematic effect on the results of the experiment. The absences or illnesses of individual pupils, however, may be considered as typical of classes in a school and their effects on the results of the experiment should be interpreted as "typical" regardless of the methods used.

Assuming that all the foregoing factors have been controlled, the observed differences in the treatment means for diets or for methods are due only partly to actual differences in the effectiveness of the treatments. They are also partly due to the effects of extraneous variables or factors, hereafter indicated as *errors* of various kinds.

BASIC TYPES OF DESIGN ERRORS

Some errors may be described as *constant errors,* or those which are present in nearly all experiments. They pertain to systematic effects which are unknown or unaccounted for and cannot be isolated from the obtained differences in treatment effects. Examples of constant errors for the illustrative problems may be such as the following.

Poultry Husbandry

1. *Diet preparation.* The diets may be composed of various dry ingredients purchased commercially and considered reliable. However, from experiment to experiment in one laboratory, or in one locality, the water used in preparation of the diets may constitute a constant error. For example, if the water is high in alkaline content or mineral content, it would affect both diets systematically but might produce a different result than if it were high in acid and low in mineral content.
2. *Weighing equipment.* The scales used in weighing out the daily portions of diets, or in weighing the various ingredients in diet preparation, may not be accurate at all weights. In this case a constant error would be present for all experiments in which these scales were used.

Education

1. *Outside activities.* In the community in which an experiment may be conducted, the out-of-school activities may have a constant effect on the experimental results. The effectiveness of the teacher with respect to teaching ability, personality, and other traits may also represent a constant error.
2. *Criterion tests.* A constant error may result from the instruments used to evaluate the two methods of instruction. They may be measuring only a limited number of the skills and concepts being taught, or be lacking in adequate reliability.

Other errors may be classified as *variable errors,* or those that might differ from experiment to experiment, or from group to

group within one experiment. These may be taken into account in the selection of the experimental design. The basic types of variable errors may be classified as sampling errors, grouping errors, and replications errors.

Type *S* (Sampling) Errors

These are the types of errors that occur due to fluctuations in random sampling. That is, the sample group assigned to treatment *A* may, by chance, contain a larger proportion of cases who can benefit by the *A* treatment than the proportion in group *B* who can benefit by the *B* treatment even though the effects of the treatments may be the same for the entire population from which the samples were drawn. For example, in the education experiment, the sample of pupils assigned to one method of instruction may, by chance, contain a larger proportion of pupils who like arithmetic, and thus learn more and make higher criterion test scores than the proportion of similar pupils in the entire population. Type *S* errors are practically always taken into account in tests of significance by the processes of random sampling and the estimates of the standard error based on sampling distributions of the subjects.

Type *G* (Group) Errors

This type of errors occurs when extraneous factors have a systematic effect on one subgroup but have no systematic effect on other subgroups. For example, one subgroup, or class of pupils, may be taught by an exceptionally good teacher whose effects are systematic on all pupils in his class but who has no systematic effect on other classes taught by other teachers. In the poultry husbandry example, the presence of a "neurotic" chick may have a systematic effect on the appetites and resulting gains in weight for the chicks in one group. If the group had been divided into subgroups, or cages, the systematic effect of the "neurotic" chick would only have been on the chicks in the cage in which he was held and would not have affected the chicks in the other cages. Type *G* errors may be taken into

account in tests of significance by randomizing the subgroups and using estimates of the standard error based on sampling distributions of subgroups.

Type *R* (Replication) Errors

When several experiments are conducted at the same time with separate samples drawn from different subpopulations of the total population with which the experiment as a whole is concerned, each of the experiments is referred to as a *replication.* These several experiments are usually but by no means always carried on by the same researcher. However, psychological experiments are often replicated by others after their results are published, as in experiments on grading essay examinations. Type *R* errors are those that occur when extraneous factors have a systematic effect on one replication but have no systematic effect on other replications. For example, the education experiment might have been conducted in a particular school whose curriculum over the past few years was particularly conducive to more favorable results for method *B* than for method *A*. Another school's curriculum might have been conducive to better result from method *A* than from method *B*. If several replications of the experiment have been conducted in several schools, the effect of Type *R* errors could have been randomized and taken into account in the estimate of the standard error used in the test of significance. The illustrations of the following three basic designs show how the Type *S*, Type *G*, and Type *R* errors are taken into account in the tests of significance.

SIMPLE RANDOMIZED DESIGN[2]

The simple-randomized design may be diagrammed for the illustrative examples as follows. Each number in the circles represents a criterion score for one case in the sample. The circles indicate that all cases under each treatment have been handled in an intact group.

[2] *Ibid.*, pp. 12–13, 47–101.

Treated Group A

11	23	32
12	24	33
15	27	33
17	29	33
18	30	44
19	30	46
19	30	48
19	30	49
20	31	51
22	31	53

Untreated Group B

12	33	39
14	33	39
16	34	40
20	34	41
22	35	52
24	35	53
28	36	54
29	37	56
30	37	57
32	38	58

The simple arithmetical mean is computed for each treatment and the magnitude of the difference between treatment means is tested for significance by use of the *t-test* for independent random samples as given in the following formula:

$$t = \frac{M_a - M_b}{\text{est'd.}\ \sigma_{(M_a - M_b)}}$$

This formula may be simplified for computational convenience to:

$$t = \frac{M_a - M_b}{\sqrt{\left(\frac{\Sigma d_a^2 + \Sigma d_b^2}{n_a + n_b - 2}\right)\left(\frac{1}{n_a} + \frac{1}{n_b}\right)}}$$

where, d_a, d_b are individual deviations from M_a and M_b, respectively, and n_a, n_b are the corresponding number of cases in each group.

The degrees of freedom for this t are $(n_a + n_b - 2)$ or 58. To be significant at the 5 percent level of confidence, the value of t would have to equal or exceed 2.00. This value is determined by referring to special tables of probability.[3] From the scores given for the two treatments above, the following measures are obtained:

$$\begin{array}{lll} M_a = 29.3 & \Sigma d_a^2 = 3970.3 & n_a = 30 \\ M_b = 35.6 & \Sigma d_b^2 = 4503.2 & n_b = 30 \end{array}$$

[3] See Table 13 in Appendix B.

By substituting these measures into the previous formula, as shown below, the value of t is found to be -2.019.

$$t = \frac{29.3 - 33.6}{\sqrt{\left(\frac{3970.3 + 4503.2}{30 + 30 - 2}\right)\left(\frac{1}{30} + \frac{1}{30}\right)}} = \frac{-6.3}{3.12} = -2.019$$

Since this value is in excess of 2.00, the difference between means may be considered statistically significant, and the hypothesis that the treatment means are equal can be rejected at the 5 percent level of confidence.

This design takes into consideration only errors of Type S. Since there is only one observation of a difference in group means there is only one observation containing a Type G error and the variability of such errors cannot be estimated. Also, Type R errors are not taken into consideration because there is only one replication. Thus, it is impossible to tell to what extent the observed treatment effect is due to Type G or Type R errors or to real differences between the treatments for the entire population.

RANDOM-REPLICATIONS DESIGN[4]

If, instead of using all cases under each treatment as intact groups, the cases were divided at random into six subgroups, with three subgroups assigned at random to each of the treatments, the design may be diagrammed as shown on page 189.

If this were a school experiment each subgroup might consist of a separate class in one school with no teacher assigned to more than one class. Any systematic effect, such as a "neurotic" pupil, or the effects of *one* teacher, would only affect one replication instead of all cases in the sample under one treatment and would, therefore, have less influence on the overall treatment means. Thus, the design would not only take into consideration Type *S* errors, but Type *G* errors as well.

If, however, the experiment were concerned with a total population consisting of a large number of subpopulations, each

[4] Lindquist, *op. cit.*, pp. 17–20, 190–202.

	Treated Group A			Untreated Group B			
	11	18	33	12	30	40	
Replication #1	12	24	33	16	32	54	$D_1 = (M_{a1} - M_{b1})$
	15	30	53	28	33	58	$= -7.9$
		32			37		
	$M_{a1} = 26.1$			$M_{b1} = 34.0$			
	17	22	31	20	34	37	
Replication #2	19	30	44	22	35	38	$D_2 = (M_{a2} - M_{b2})$
	19	30	46	29	35	53	$= -5.0$
		31			36		
	$M_{a2} = 28.9$			$M_{b2} = 33.9$			
	19	27	48	14	34	52	
Replication #3	20	29	49	24	39	56	$D_3 = (M_{a3} - M_{b3})$
	23	30	51	33	39	57	$= -6.0$
		33			41		
	$M_{a3} = 32.9$			$M_{b3} = 38.9$			
General Means	$M_a = 29.3$			$M_b = 35.6$			$\overline{D} = (M_a - M_b) = 1/3(D_1 + D_2 + D_3) = -6.3$

replication might consist of two or more classes in each of three schools. Any systematic effect of any given school curriculum would then affect only one replication instead of all cases in the total sample. The schools should be selected at random, pupils assigned to the classes at random, and the treatments assigned at random to the classes. Thus, the design would include Type *R* errors as well as Type *S* and Type *G* errors.

The effect of treatments is determined in this design by the following *t*-test:

$$t = \frac{M_a - M_b}{\text{est'd. } \sigma_{(M_a - M_b)}}$$

This formula may be simplified for computational convenience to:

$$t = \frac{\overline{D}}{\sqrt{\dfrac{\Sigma(D_i - \overline{D})^2}{n(n-1)}}}$$

where, $\overline{D} = (M_a - M_b) = \frac{1}{3}(D_1 + D_2 + D_3)$
D_i = each of the replication D's, or D_1, D_2, and D_3
n = the number of replications, or, in this case, 3

The degrees of freedom for this t are $(n - 1)$, or 2. To be significant at the 5 percent level of confidence, the value of t would have to equal or exceed 4.303. From the scores given in the above diagram, the following measures are obtained:

$$M_a = 29.3 \qquad M_b = 35.6 \qquad \overline{D} = -6.3 \qquad n = 3$$
$$(D_1 - \overline{D})^2 = 2.56 \qquad (D_2 - \overline{D})^2 = 1.69 \qquad (D_3 - \overline{D})^2 = 0.09$$

By substituting these values into the previous formula, as shown, the value of t is found to be -6.678.

$$t = \frac{-6.3}{\sqrt{\dfrac{2.56 + 1.69 + 0.09}{3(3 - 1)}}} = -6.678$$

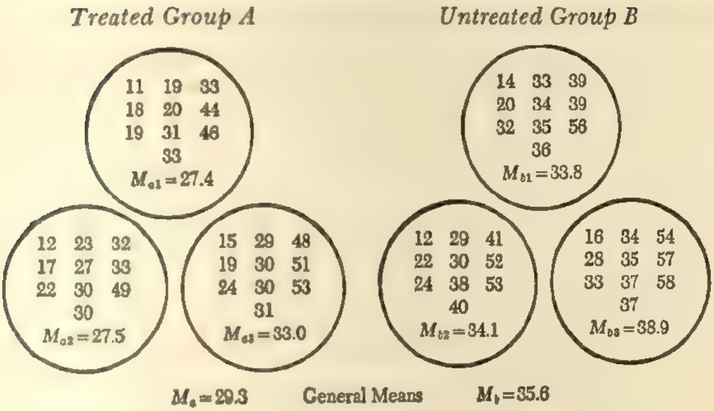
Since this value is in excess of 4.303, the difference in treatment means may be considered greater than could be attributed to chance fluctuations in sampling. An inference that there are real differences in the treatments can be drawn with more assurance than in the previous design since this design has taken into account differences in the results that might be due to systematic effects of subgroups and has freed them from the treatments effects.

GROUPS-WITHIN-TREATMENTS DESIGN[5]

The groups-within-treatments design may be used to establish generalizations about a population consisting of a large number of subpopulations when it is not possible to use the random-replications design. The design may be diagrammed as shown on page 191.

In this design the pupils selected at random who are randomly assigned to each class (see small circles) are assumed to be a random sample of all pupils who might have been in that class. The design regards each general treatment mean as an average of a random sample of class means rather than a random

[5] *Ibid.*, pp. 23–25, 172–187.



sample of individual pupils. Since the classes under each treatment are a random sample of all classes in the population, the Type *S*, Type *G*, and Type *R* errors in the means for these classes may be regarded as a simple random sample from the distribution of such errors that would be found for all classes if all used the same method.

The effect of treatments is determined in this design by the following *t*-test:

$$t = \frac{M_a - M_b}{\sqrt{\text{est'd}\,\sigma_{M_a}{}^2 + \text{est'd}\,\sigma_{M_b}{}^2}}$$

This formula may also be simplified for computational convenience to:

$$t = \frac{M_a - M_b}{\sqrt{\dfrac{\Sigma(M_{ai} - M_a)^2}{c_a(c_a - 1)} + \dfrac{\Sigma(M_{bi} - M_b)^2}{c_b(c_b - 1)}}}$$

where, $M_a, M_b = \frac{1}{3}(M_{a1} + M_{a2} + M_{a3})$; and similarly for M_b

$M_{ai}, M_{bi} =$ each of the training group means, or M_{a1}, M_{a2} and M_{a3}, and similarly for M_{b1}

$c_a, c_b =$ the number of training groups in each treatment or 3.

The degrees of freedom for this t are $(c_a - 1) + (c_b - 1)$, or 4. To be significant at the 5 percent level of confidence, the value

of t would have to equal or exceed 2.776. From the scores given in the above diagram, the following measures are obtained:

$$M_a = 29.3 \qquad M_b = 35.6 \qquad c_a = 3 \qquad c_b = 3$$

$$\Sigma(M_{ai} = M_a)^2 = [(1.9)^2 + (1.8)^2 + (3.7)^2] = 20.54$$
$$\Sigma(M_{bi} - M_b)^2 = [(1.8)^2 + (1.5)^2 + (3.3)^2] = 16.38$$

By substituting these values into the previous formula, as shown, the value of t is found to be −2.540.

$$t = \frac{29.3 - 35.6}{\sqrt{\dfrac{20.54}{3(3-1)} + \dfrac{16.38}{3(3-1)}}} = -2.540$$

Since this value is less than 2.776, the obtained difference in treatment means can *not* be considered greater than could be attributed to chance factors, and the hypothesis of equal means can *not* be rejected. As this design takes into consideration all three types of errors (S, G, and R), it is apparent that the overall difference in treatment means is not large enough to account for real differences in excess of those that may be accounted for by chance factors when samples of this size are used.

If, however, the sample had been double the size of the present one, so that there were six training groups instead of three, and if they had had the same means and the deviations of the group means from the general means were alike, the following measures would have been obtained:

$$\Sigma(M_{ai} - M_a)^2 = 41.08, \ \Sigma(M_{bi} - M_b)^2 = 32.76, \ c = 6$$

The degrees of freedom would have been increased to ten, and the value of t would only have to equal or exceed 2.228. By substituting these new values into the formula, the value of t would have been found to be −4.018. This value would have been significant at the 5 percent level of confidence and the difference in treatment means would have been considered greater than could have been expected by chance alone. This illustration points out the danger of using samples that are too small in which the chance effects of sampling are likely to obscure real

differences that would have appeared if the sampling had been adequate.

It should be noted that each of the three designs presented made use of the criterion measures for a total of 60 individuals. By the way in which this number of individuals is sampled it may be seen that different types of error can be taken into account in testing the significance of difference between treatment groups. Thus, the method of analysis should be determined prior to selecting the sample to insure that the sampling procedures will be most effective.

OTHER DESIGNS

The three basic designs previously illustrated involved only two treatments and can be tested for significance of difference by use of the *t*-test. When comparisons are to be made simultaneously among several treatments, different analytical procedures and different tests of significance are employed, although the basic designs remain very much the same. The following are some of the more complex designs, involving analysis of variance techniques:

Treatments-by-levels and treatments-by-subjects
Factorial experiments
Split-plot designs
Three dimensional
Higher-dimensional
Complete and incomplete factorial designs
Latin square and Greco-Latin square
Incomplete Latin squares (Youden squares)
Lattice squares and cubic lattice squares
Randomized block and incomplete block designs

The techniques of applying these designs and the statistical procedures used are discussed in several of the textbooks listed at the end of this chapter.

TREATMENT OF EXTRANEOUS VARIABLES

In all the above designs, the major problem is concerned with the treatment of extraneous variables. In educational research experimental methods of instruction, for example, are influ-

enced by the effects of teachers, backgrounds of pupils, environmental effects, and so forth, which are included in the overall measures of differences in methods. An *ideal* technique would involve the *elimination* of all extraneous variables. However, in most practical situations, it is impossible to eliminate them, so they have to be treated through other processes, such as equalization, counterbalancing, or randomization.

Equalization

In equalization every case in every group is treated identically with every other case, or each case in an experimental group is matched by a corresponding case in one or more control groups and treated with respect to the extraneous variable under consideration. The process of equating groups to take the extraneous variable into account has been made on various bases which may be indicated by the following comparisons:

Chance versus actual measurement.

One trait versus several traits.

The average versus both the average and variability of traits.

Composite scores for all traits versus scores for each trait separately.

Average and variability for the group as a whole versus the pairing of individuals.[6]

Chance equating is valid only when very large numbers of cases are used and when groups are selected "at random." Groups may be alike in one trait but differ in another and should be equated with respect to all the traits that may influence the experimental factors. Where these factors have varying influences upon persons of different "levels," it is desirable to have the two groups of equal variability as well as of equal average value. When the composite of scores on several traits is used, the score for each trait should be properly "weighted" before being combined into a composite score. This is best accomplished through the use of "standard scores" which take into account the variability of all cases in a group. Probably the best

[6] Harold H. Abelson, *The Art of Educational Research*. Yonkers, N.Y.: World Book, 1953, pp. 145–148.

method is to pair each person in one group with a corresponding person in another group on the trait considered or on a composite of traits. If individuals are paired, the average scores and the variability of the groups will tend to be equalized. However, it is often difficult to pair on even two or three variables without rapidly losing cases that can not be paired and thus reducing the effective size of a sample.

Counterbalancing

A method of systematic counterbalancing is employed where equalization is not possible so that the total effects of the variables in each group are equalized even though these effects may vary within each group. This method is sometimes referred to as the "rotation-group technique."[7] The process of counterbalancing may be illustrated by a classical example of determining whether a subject has a better sensitivity to various tones with his left ear or with his right ear. His threshold of sensitivity might be determined by making a test six times with each ear. Suppose the tests were made six times on the left ear first, and then six times on the right ear. In listening for certain tones, the subject may reflect a difference that is due to an increased sensitivity to tones through testing the right ear. To avoid this "practice" effect, the series of tests could be counterbalanced by alternating the left and right ears, as L, R, L, R, L, R—R, L, R, L, R, L.

In determining the effectiveness of various methods of instruction with training groups, to cite another example, there may be a carry over of effects from one instructional procedure to another, when the different methods are used successively on the same classes, which is further complicated by a general increase in pupil achievement and maturation due to the amount of time involved in the course of the experiment. If a minimum of six classes were used in an experiment involving three different methods of instruction (A, B, and C) over a period of time, these effects would tend to be equalized by using the coun-

[7] Carter V. Good and Douglas E. Scates, *Methods of Research*. New York: Appleton-Century-Crofts, 1954, pp. 706–707.

terbalancing technique. This technique is one in which the order of using the different methods with each class could be as follows:

Class I	A, B, C	Class IV	A, C, B
Class II	B, A, C	Class V	B, C, A
Class III	C, A, B	Class VI	C, B, A

In this manner every combination of effect of one method on another has been counterbalanced. Naturally it would be desirable to have more than one class for each of these sequences to be able to take into account systematic differences among classes that are not due to differences in the methods themselves.

Randomization

The method of randomization is appropriate where it is impossible to use the equalization technique or the method of counterbalancing. This method may also be used to provide an overall randomization where the particular extraneous variable is undetermined, or it may be used in combination with the method of equalization where the extraneous variable can be measured to provide a basis for matching experimental and control cases. In this method the subjects or treatments are assigned to the experimental or control groups in some purely random fashion following the procedures discussed under the topic Random Sampling, in Chapter III.

SELECTED REFERENCES

Chapin, F. Stuart. "Some Problems in Field Interviews When Using the Control Group Technique in Studies in the Community," *American Sociological Review, 8*:63–68, 1943.

Cox, D. R. *Planning of Experiments.* New York: Wiley, 1958. 308 pp. (A good basic reference on statistical designs to supplement courses in statistics.)

Edwards, Allen L. *Experimental Design in Psychological Research.* Rev. ed. New York: Holt, Rinehart, and Winston, 1960. 398 pp.

Fisher, R. A. *The Design of Experiments.* 6th ed. New York: Stechert-Hafner, 1951. 244 pp. (See index for special topics.)

Gage, N. L. (Ed.). *Handbook of Research on Teaching.* Chicago: Rand McNally, 1963. (See Part III.)

Johnson, Palmer O. *Statistical Methods in Research.* Englewood Cliffs, N.J.: Prentice-Hall, 1949. 377 pp. (Read pp. 276–358.)

Kempthorne, Oscar. *The Design and Analysis of Experiments.* New York: Wiley, 1952. 631 pp. (See index for desired topics.)

Lindquist, Everet F. *Design and Analysis of Experiments in Psychology and Education.* Boston: Houghton Mifflin, 1953. 393 pp. (An exceptionally straightforward discussion of research designs and analyses.)

Ray, William S. *An Introduction to Experimental Design.* New York: Macmillan, 1960. 254 pp.

Van Dalen, Deobold, and Meyer, William J. *Understanding Educational Research.* New York: McGraw-Hill, 1962. (See chap. 11.)

Winer, B. J. *Statistical Principles in Experimental Design.* New York: McGraw-Hill, 1962. 672 pp.

CHAPTER IX

Scaling Problems and Techniques

Rating Scales
- The continuum
- Examples of rating scales
- Number of units in a scale
- Rules for constructing graphic and numerical scales
- The raters
- Pooled ratings

Rank-Order Scales

Social Distance Scales

Sociometric Analysis
- Sociometric matrix
- Sociogram
- Sociograph

Determination of Scale Values
- Method of equal-appearing intervals
- Method of paired comparisons
- Rank order method
- Other techniques

Research in the human and behavioral sciences is based primarily upon qualitative facts dealing with how people feel, what they have experienced and remembered, what their emotions and motives are like, and the reasons for acting as they

do. The problem of converting a series of qualitative facts (opinions, impressions, convictions, attitudes, interests, and the like) into a quantitative series more readily amenable to analysis and interpretation has been one of the most perplexing problems to researchers. For example, assume that it is required for the solution of some specific problem to make comparisons, or to determine differences, among groups of people on the basis of their socioeconomic status. In making such an analysis one may have to take into consideration their material possessions, the social group in which they move, their occupations, their incomes, and several other factors. The difficulties of analysis occur when it is found that some individuals who are "high" on one attribute and "low" on a second attribute appear to have the same status as other individuals who are "low" on the former and "high" on the latter. A similar difficulty is encountered when one attempts to determine varying degrees of favorable or unfavorable attitudes toward some particular phenomenon.

It is obvious that differentiations among these types of qualitative facts are possible for extreme cases. That is, it is easy to make a fairly valid qualitative judgment that one person is extremely favorable to a "single salary schedule for men and women teachers," for example, while another person is extremely unfavorable to it. However, difficulties are met when one attempts to differentiate among all those individuals who indicate that that they are favorable to a single salary schedule. It is not easy to ascertain how much more favorable one individual is than another, and this difficulty increases as the differences in their attitudes become less and less. Thus, it is necessary to place these attributes in an ordered series from one extreme to the opposite (favorable to unfavorable), to increase the number of ordered categories to yield smaller and smaller differences between adjacent categories, and to convert them into variables that can be expressed in numerical fashion.

In handling such problems it is necessary, first of all, to assume that in the variable under consideration there is a *continuum* of values; secondly, to assume that this continuum can

be expressed in *quantitative terms*; thirdly, it must be assumed that *representativeness* of the continuum can be achieved by a particular set of quantitative values. If the qualitative facts under consideration could be arranged into a quantitative series that had a zero point and equality of units of measurement that were perfectly reliable and valid, the statistical procedures discussed in previous chapters could be applied with little difficulty. There is also some question as to whether scales of qualitative facts can be thought of as employing *cardinal numbers* such that the units of measurement could be added, subtracted, multiplied, or divided. However, it is generally agreed that qualitative facts can be expressed in *ordinal units* and that it is not always necessary for a scale to possess either a zero point or absolutely equal units to be useful.

The methods and techniques of scaling qualitative facts, as discussed in this chapter, may be classified into the following general categories: rating scales, ranking scales, social distance scales, sociometric measurement, and the determination of scale values.

RATING SCALES

Rating is a term applied to the expression of opinion or judgment regarding some situation, object, person, etc. These opinions are usually expressed on a scale or by categories of values, either quantitatively or qualitatively. For example, a teacher in assigning a mark or grade to a pupil is applying a rating scale of proficiency of some sort to that pupil; or an individual in filling out a recommendation form for another person for a teacher's placement bureau is frequently using some sort of a rating scale. Rating scales are probably one of the more commonly used forms for scaling traits and attributes.

The techniques of rating should take into account some psychological continuum that can be defined as unequivocally as possible, certain definitive aspects of the continuum that can aid a judge to evaluate the "subjects" and to place them on the continuum, and the reliability and validity of the opinions of the judges.

The Continuum

Most rating scales have a continuum that may be defined by descriptive phrases to which numerical or letter values are assigned. Several kinds of scales are used, such as numerical ratings, graphic ratings, and descriptive ratings. In all rating scales explicit directions should be provided which indicate the nature of the continuum and the scale points to be used in rating the subjects. Care should be taken to avoid using ambiguous descriptive terms, or terms such that the raters are unable to distinguish accurately between units because of differing personal beliefs and values. Such terms as "noncoöperative," "antagonistic," "belligerent," "rebellious," and so forth, have different meanings for different raters.

Examples of Rating Scales

The *numerical rating scale* is usually combined with some descriptive phrases of a trait which may be judged according to a number of steps or units. For example:

Directions: The items on this scale are to be rated on the basis of 1 for unsatisfactory or failure, 2 for inferior, 3 for average, 4 for better than average, and 5 for outstanding or superior. Circle the number that corresponds to your rating of this person for each of the following items.

1. Showing others how to do things	1	2	3	4	5
2. Clean and attractive personal appearance	1	2	3	4	5
3. Particular about his (her) work	1	2	3	4	5
(continued in like manner)					

A *graphic rating scale* usually consists of a line on which a rater indicates by a check mark the point most nearly representing his rating of an individual. An example of a graphic rating scale is as follows:

Behavior Rating Scale

Directions: Make a cross (X) on the line for each of tne following items to indicate how you would rate the subject named.

Name: ________________ Date: __________ Rater: __________

1. Enthusiasm

Indifferent	Rarely shows enthusiasm	Sometimes enthusiastic	Usually has pep and vigor	Works with great zeal

2. Integrity

Positively dishonest	Honesty doubtful	Usually reliable	Apparently honest	Unquestionably honest

(continued in like manner)

The *descriptive rating scale* usually presents a number of phrases descriptive of varying degrees of a trait or characteristic. The following excerpt of a record developed by the San Diego Public Schools is illustrative.

Behavior Rating Observation [1]

To understand behavior, it is important to observe the pupil's reactions on the playground, in the neighborhood, and at home. Check words or phrases that describe the behavior of the pupil as you have observed it. Please feel free to individualize the report as much as possible by adding descriptive comments. If you know of reasons for the conditions you check, please jot them down at the right of your answers.

IS THIS PUPIL PHYSICALLY STRONG?

...... Is strong and active.
...... Seldom tires.
...... Has ordinary endurance.
...... Is listless, easily fatigued.

DOES HE HAVE GOOD WORK HABITS?

...... Completes what he starts.
...... Is able to evaluate his work.
...... Capable of sustained attention.
...... Is easily discouraged.
...... Seldom completes the job.
...... Easily distracted.

[1] "Behavior Observation Record." Mimeographed check list. San Diego Public Schools, 1949, cited in *Evaluating Pupil Progress*, Bulletin of the California State Department of Education (April, 1952), Vol. XXI, No. 6, p. 127.

...... Needs urging to stay with a task.
(continued in like manner)

The following form may be used for the judgment of subjects with respect to their qualifications for a college teaching position based upon a judge's knowledge of the subjects. This form is somewhat of a combination of the first and third illustrations, above.

ESTIMATE OF NOMINEE'S QUALIFICATIONS BASED ON CLASSROOM RECORD

Rate the individual named on each of the characteristics listed below by placing a cross (X) in the proper column. The scale follows: *A*, Superior; *B*, Above Average; *C*, Average; *D*, Below Average; and *F*, Unsatisfactory.

	A	*B*	*C*	*D*	*F*
1. Character					
2. Class attitude					
3. Cooperation					
4. Emotional poise					
5. Health and vitality					
6. Industry and dependability					
7. Intelligence and adaptability					
8. Knowledge of subject matter					
9. Personal appearance					
10. Speech					

NUMBER OF UNITS IN A SCALE

Deciding the number of units to use in a scale is largely an empirical matter. The above illustrations of numerical and graphic rating scales have used five points on the "continuum." If fewer steps were used, the scale would obviously be a coarse one and would have little meaning. If a large number of units were used, the discriminations among them would become increasingly more difficult and the scale would tend to become less reliable. If a trait is to be analyzed into three categories it is desirable to use a five-point scale since judges are inclined to hesitate to give extreme judgments and tend to displace indi-

viduals in the direction of the mean of the total group. With a three-point scale, then, there would be a tendency to have most of the ratings fall into only one category. If more than a five-point scale is used it is necessary to have very definitive descriptions of each point or step being used.

Rules for Contructing Graphic and Numerical Scales

Since graphic scales are used commonly, it is desirable to present here a compilation of rules for their construction. It should be remembered that the application of these rules should be made in terms of the type of information desired and the manner in which it is to be analyzed. The following rules have been presented by Guilford.

1. Each trait should occupy a page by itself. This rule is rarely observed. When numbers of individuals are to be rated, it is far better that all of them be rated in one trait before going on to the next.
2. The line should be at least 5 inches long but not much longer, so that it can be easily grasped as a whole.
3. The line should have no breaks or divisions. A broken line may suggest a discontinuous variable.
4. The "good" and "poor" ends should be alternated in random order so as to avoid a constant motor tendency to check at one side of the page.
5. Introduce each trait with a question to which the rating gives an answer, that is, "How has he responded when praised?"
6. Use three or five descriptive adjectives—two extremes and one or three intermediates.
7. The descriptive phrases should be in small type with considerable white space between them.
8. Only universally understood descriptive terms should be used, avoiding slang or other colloquial expressions.
9. Decide beforehand upon the probable extremes of ability (or of the trait) to be found in the group or groups in which the scale is to be used.
10. The end phrases should not be so extreme in meaning as to be avoided by the raters.

11. Have the extreme phrases set flush with the ends of the lines.
12. The average or neutral phrase should be at the center of the line.
13. Descriptive phrases need not be evenly spaced. The meaning of the intermediate ones should be nearer the middle one than the extremes.
14. In the scoring use a stencil which divides each line into several sections to which numerical values are assigned.
15. The divisions of the scoring stencil need not be equal; they may be made to conform to the distribution of ratings.
16. Do not require any less distinctions in rating than are used in scoring. If anything, the scoring units may be less than the rating units.[2]

The Raters

The effectiveness of measurement by rating methods requires not only specificity and comprehensiveness of definition of the traits rated, but also depends upon the ability of the raters to discriminate reliably. In general, pooled ratings increase the accuracy of any rating scale. In addition, it is necessary to select raters representative of the population and expert with respect to the traits being rated. They should be selected on the basis of expertness in relation to the continuum being used in so far as it is possible.

Pooled Ratings

Ratings may be made in an "order of merit" by means of some criterion values ranging, for example, from 1 to 10 whereby each subject is given a value somewhere within that range but in which it is not necessary to use all values. This procedure may be illustrated by the following table showing the ratings assigned by six judges (teachers in a school) to eight pupils with respect to their "integrity." The degrees of integrity, ranging from "unquestionably honest" to "usually reliable" to "positively dishonest" were recorded on a continuum having numer-

[2] J. P. Guilford, *Psychometric Methods*. New York: McGraw-Hill, 1936, pp. 271–272.

	Pupils							
Judges	1	2	3	4	5	6	7	8
1	10	9	6	5	4	7	1	3
2	9	7	8	6	4	1	3	5
3	9	7	8	6	5	3	4	1
4	7	10	8	9	3	5	3	4
5	8	8	7	4	6	3	5	2
6	10	8	8	7	5	5	3	1
Sum	53	49	45	37	27	24	19	16
Mean	8.8	8.2	7.5	6.2	4.5	4.0	3.2	2.7
Median	9.0	8.0	8.0	6.0	4.5	4.0	3.0	2.5

ical values ranging from a high of 10 to a low of 1. Either the mean or the median values at the bottom of the table may be used as "pooled" ratings for the final ratings of these eight pupils. These pooled values could subsequently be converted into actual scale values by means of techniques for determining scale values discussed in a later section of this chapter.

RANK-ORDER SCALES

Rank-order scaling is quite similar to rating techniques with respect to the selection and use of judges as raters. However, it differs somewhat with respect to the continuum in that the scaling of judgments is made by comparing subjects with each other rather than judging them on a trait scale. The members of a group are compared on certain qualities commonly shared in varying degrees, and there are no numerical units used other than the numbers indicating the serial position of each member in the group. Rank-order methods may be used advantageously in the evaluation of samples of handwriting, determining what factors occur most frequently that affect teacher morale, the ethical judgments of elementary school children about various social offenses, and other problems in which an order of values is considered important.

In using the rank-order method, the subject (products, traits, attributes, or any other phenomena being ranked) are placed in a special order by each of a number of judges. In general, the more judges used, the more reliable the rank order will be. The

mean or median rank is then computed for all judges. These mean or median rankings are then considered as the final rank order representing the pooled judgments of a number of judgments. They do not necessarily represent scale values in themselves, but they may be used as approximations of linear scale values when the distribution of the phenomenon being ranked approaches the rectangular type. When the distribution is bell-shaped, or approaches the normal type of distribution, the ranks cannot represent scale values.

The ranking of subjects may be carried out in a "man-to-man" order, whereby the total number of ranks to be used is equal to the number of subjects being ranked. That is, if 12 pupils are to be placed in rank order, their ranks would range from 1 to 12 as integral values, except for the cases of ties in which the ranks the tied subjects would occupy are averaged and the average value is given to each. (This is the same process as shown in the computation of rank-order correlation coefficients in Appendix B.) When the ranking of subjects is made by a number of judges, their pooled ranks provide a final ranking based on either the mean or median values. These pooled ranks could subsequently be converted to scale values by techniques to be discussed later.

SOCIAL DISTANCE SCALES

Social distance refers to the "grades and degrees of understanding and intimacy which characterize presocial and social relations generally."[3] The concept of social distance applies to an individual, a social group, or a value in terms of some continuum ranging from close, warm, intimate acceptance at one extreme to active dislike, hostility, and rejection at the other extreme. In this process a subject is asked to state whether he would accept, for example, a certain kind of person as a citizen, as a neighbor, or as a member of his family. When the process is carried out with a large group of individuals, it is possible to determine what kinds of persons are acceptable to the group,

[3] Robert E. Park, "The Concept of Social Distance," *Journal of Applied Sociology* (1902), 8:339–344.

the order of their acceptability, and other information of sociological value.

Pioneer work was carried out by Bogardus[4] in the development of scales for studying ethnic problems. The following is an illustration of a Bogardus-type scale usable for determination of immigration and race attitudes.

Directions: According to my first feeling reactions I would admit members of each race or nationality (as a class and not the best I have known, nor the worst members) to one or more of the classifications which I have circled according to the following:

1. To close kinship by marriage.
2. To my club as personal chums.
3. To my street as neighbors.
4. To employment in my occupation.
5. To citizenship in my country.
6. As visitors only to my country.
7. Would exclude from my country.

ETHNIC GROUPS	CLASSIFICATIONS						
English	1	2	3	4	5	6	7
Swedes	1	2	3	4	5	6	7
Poles	1	2	3	4	5	6	7
Chinese	1	2	3	4	5	6	7
Negroes	1	2	3	4	5	6	7
(etc.)							

After a group of respondents has recorded its reactions to the ethnic groups included in the scale, frequencies of responses are tabulated for each classification. These may be converted to percentages of responses for all ethnic groups on each attribute or classification, or by percentages of responses for all classifications on each ethnic group. Thus, such tabulations would represent not only one scale but a number of scales. The percentages can be used to determine the rank order of groups,

[4] Emory S. Bogardus, "Social Distance and Its Practical Implications," *Sociology and Social Research* (1953), *17*:265–271.

but methods of quantifying the scales to show the relative acceptability of one group over another require more complex mathematical constructs.

This technique has flexibility for use in many applications but it is limited in its precision of measurement. It is difficult to test the reliability of the scale except by the test-retest method. While the scale does provide an ordering along some sort of continuum, the assumption that each point is necessarily beyond the preceding one is difficult to justify, and it is almost impossible to make the necessary assumption that there is an equal amount of social distance between each point on the scale and the succeeding one. For further details of the development and use of social distance scales the reader is referred to the selected references at the end of the chapter.

SOCIOMETRIC ANALYSIS

Sociometric analysis is used to determine the structure of relationships at a given time among members of a given group. It may be used advantageously to judge the social climate of a classroom or the social status of individual children. It is possible for one to gain considerable information about individuals' social relationships through observation, but sociometric techniques may be used to reveal how individuals would *like* to associate and how their wishes compare with the feelings of other individuals toward them.

In making a sociometric analysis, it is necessary first of all to obtain from each individual in a group his choices and order of choices of others in the group for a specific situation. For example, a situation may be structured on the basis of asking pupils in a classroom, "With whom would you like to work and play the best?" Each pupil might be asked to write on a slip of paper the names of other pupils in the class with whom he would like to work and play as first, second, and third choices. These slips would then be collected and the information supplied would be recorded on a two-way table or *sociometric matrix,* as shown in Table 3.

As far as sociometric analyses are concerned, more complete and more valid analyses may be made if pupils were asked to nominate not only the ones they like the best, but also the ones they like the least. However, there is considerable question about the desirability of having pupils identify and report to others negative assessments of their peers. Thus, the analysis illustrated here only includes the positive assessments of a group of fifth-grade pupils.

Sociometric Matrix

The sociometric matrix should include the names of all pupils in the class, listed alphabetically. These are then numbered in sequence across the top and down the left side of the two-way table. From the slip of each child, record the number of his or her choice of each of three other children. For example, Alice (pupil #1) indicated her choices of Sue, George, and Mark as first, second, and third choices, respectively. These are recorded in the first row of Table 3. At the bottom of the table are four rows for indicating the frequency of each choice for each individual and the total number of times he was chosen.

By using the frequency of choices at the bottom of the sociometric matrix, one may determine who the more popular individuals are in a class and those who are somewhat neglected or isolated from the others. The frequencies in this example indicate that Bill, JoAnn, and Lloyd are the more popular pupils in the class, and that Larry, Lois, and Nelda are the less popular.

Sociogram

Additional information may be revealed by making a *sociogram,* or pictorial illustration of the interactions among the pupils. The sociogram in Figure 21 represents the data in Table 3. It could have been constructed by simply showing only the first, or the first and second choices, but here all three choices are represented. As the number of choices increases, however, the complexity of the sociogram increases and it soon becomes exceedingly difficult to interpret.

One of the more popular techniques for constructing a

TABLE 3. Sociometric Matrix Showing Who Chooses Whom to Work and Play With in an Elementary School Class

Chooser		Pupils Chosen 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Alice	1							2										3						1
Alex	2				2						1								3					
Betty	3									3					1		2							
Bill	4										2		3						1					
Evelyn	5										3				2							1		
Floyd	6	2													1				3					
George	7											1						3						2
Glenn	8									2											3	1		
Jerry	9			3	2														1					
JoAnn	10				3								1		2									
June	11							1					3					2						
Karen	12				1						2				3									
Larry	13	3				2	1																	
Lloyd	14		2		3						1													
Lois	15				3	2					1													
Lottie	16		2								1				3									
Mark	17							1				2												3
Mary	18				1					3					2									
Nelda	19		2										3					1						
Olaf	20		3		2																		1	
Pearl	21				1			2										3						
Sarah	22						3		2								1							
Sue	23		2		1						3													
First					4		1	2			4	1	1		2		1	1	2			2	1	1
Second		1	4		3	2		2	1	1	2	1			3		1	1						1
Third		1	1	1	3		1			2	2		3		2			3	2		1			1
Total		2	5	1	10	2	2	4	1	3	8	2	4	0	7	0	2	5	4	0	1	2	1	3

sociogram is that of Northway's[5] target diagram which contains four concentric circles similar to an archery target. The concentric circles are drawn equal distances apart, and the pupils

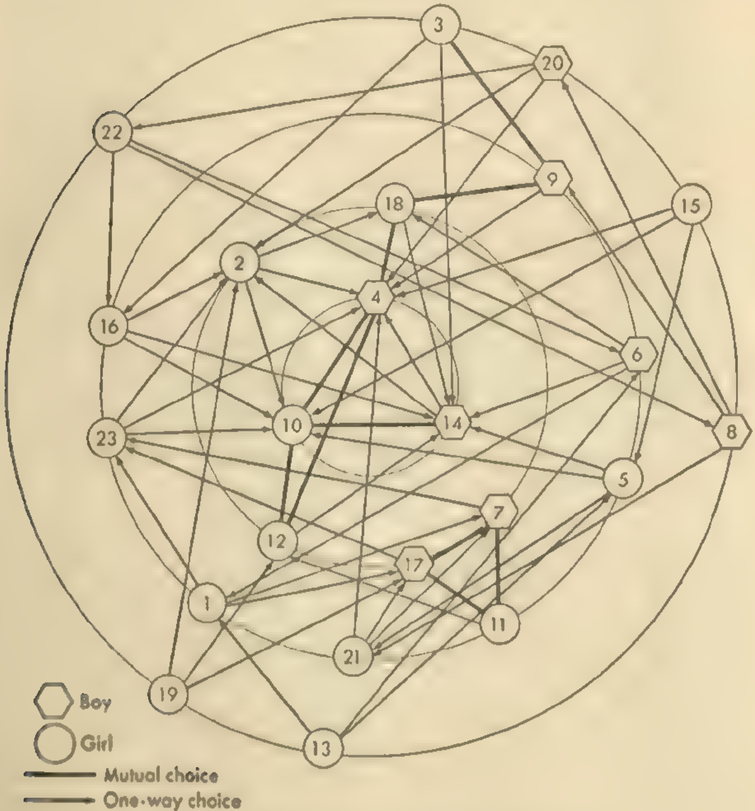


FIGURE 21. Sociogram of the Data in Table 3.

receiving the larger number of choices (stars) are placed in the smallest circle in the center of the diagram; pupils receiving only one choice (neglectees) or none (isolates) are placed in the outer ring of the diagram; pupils between these extremes are

[5] Northway, M. L., "A Method for Depicting Social Relationships Obtained by Sociometric Testing," *Sociometry*, 3:144–150, 1940.

placed in the two inner rings indicating above or below average status depending on the number of choices received. Sometimes sociograms are constructed with boys on one side of a vertical line passing through the center of the diagram and with girls on the other side.

In this sociogram, boys are indicated by hexagonal figures and girls by circles, although other shaped figures are frequently used for this purpose, and their corresponding sociometric matrix number is inserted for identification. The heavy lines connecting pupils 4, 10, 12, etc., indicate *mutual* choices, or children who have chosen each other. The lighter lines indicate one-way choices with the arrows pointing to the pupils chosen.

It is usually asumed that the pupils who are frequently chosen by other pupils for close association have a high degree of social acceptability to their fellow pupils. In this particular example, it can be seen that pupils numbered 4, 10, and 14 are near the center and represent the more popular pupils. Individuals 22, 20, 3, and 8 are evidently less liked (neglectees), while 15, 13, and 19 appear to be "isolates."

One of the major difficulties in constructing sociograms is that many different-appearing configurations of interactions may be drawn from the same sociometric matrix and somewhat different interpretations might be made from an inspection of the various mazes of lines so constructed.

Sociograph

A more reproducible graphical representation of the sociometric data, as suggested by Clark and McGuire,[6] is called a *sociograph*. It not only shows the arrangement of pupils systematically in discernible cliques and other subgroups, but also indicates cleavages in the group structure. Figure 22 is a sociograph of the previous data.

Sociographs may be constructed in different ways so as to be reproducible, each of which may be based upon some logical sequence of persons chosen by their peers. For example, one

[6] Rodney A. Clark and Carson McGuire, "Sociographic Analysis of Sociometric Valuations," *Child Development, 23* (2): 129–140, June, 1952.

TABLE 4. Determination of Sociographic Sequence

Chooser	Working List Chosen	Extra Times Chosen	Sociographic Sequence
4	18, 10, 12	x x x x x x	4
18	4, 14, 9	x x	18
10	12, 14, 4	x x x x	10
12	4, 10, 14	x	14
14	10, 2, 4	x x x x	2
9	18, 4, 3	x	12
2	10, 4, 18	x	9
3	14, 16, 9		3
16	10, 2, 14		16
		Cleavage	
17	7, 11, 23	x x	17
7	11, 23, 17	x	7
11	7, 17, 12	x	11
23	4, 2, 10	x	23
		Cleavage	
21	4, 7, 17	x x	21
6	14, 1, 18	x x	6
1	23, 7, 17	x	1
5	21, 14, 10	x	5
22	16, 8, 6		22
8	21, 9, 20		8
20	22, 4, 2		20
19	17, 2, 12		19
13	6, 5, 1		13
15	10, 5, 4		15

The sociograph is then constructed by listing persons according to their order in the sociographic sequence along the left-hand side and across the top of a sheet, as indicated in Figure 22. A diagonal line is drawn on the sociograph through the positions corresponding to the same name vertically and horizontally. Opposite each person listed, an *X* is placed to indicate a mutual choice, and a slash-mark to indicate a one-way choice. For example, Bill (No. 4) chose Mary (No. 18), JoAnn (No. 10), and Karen (No. 12), but Mary and JoAnn also chose Bill as indicated by the *X*'s.

The cleavage lines are drawn horizontally and vertically on

the sociograph following the last person in each sequence group as shown in Table 4. Sometimes these cleavage lines are used to indicate different sociographic levels (I, II, etc.) and may be used in the development of indexes of peer status. By means of the sociograph, cliques, clusters, or other subgroups are easily discernible.

For more detailed information in the construction of sociograms and sociographs, and for more detailed interpretations of these and other sociometric procedures, the reader may refer to materials published in *Sociometry;* and in reference books by Goode and Hatt; Gronlund; Jennings; Loomis and Pepinsky; Seltiz; and Young, as indicated in the selected references at the end of the chapter.

DETERMINATION OF SCALE VALUES

Scaling techniques are an attempt to transform a series of qualitative facts into a series of quantitative facts, or variables, so that processes of measurement and statistical analysis can be applied. Of the four types of scales (nominal, ordinal, interval, and ratio), the determination of scale values is concerned primarily with the interval and ratio types of scales.

The simplest type of scale, referred to as the "nominal," consists of two or more categories into which objects are classified. The only requirement for a nominal scale is that the categories be different from each other, but they do not need to represent any greater or lesser degree of the characteristic being categorized. Numbers may be assigned to these categories, but they are only for identifying the categories and have no mathematical relationship to each other. The only statistical use of the nominal scale is in those operations appropriate to counting, such as the number of cases, the mode, and the contingency correlation.

A more definitive scale, which reveals the relative position of objects with regard to some characteristic, is the "ordinal" scale. The requirements for it are that objects can be categorized (as in the nominal scale) and be placed in some order. This means that the various objects can be determined empirically as greater than, equal to, or less than each other, but that "how

much greater" or "how much less" cannot be stated. The statistics that can be applied to data scaled in this manner are limited to the determination of medians, percentiles, and the types of correlation coefficients whose underlying assumptions do not include equal units in the measuring instrument.

A better scale is one which includes, in addition to the properties of the nominal and ordinal scales, the characteristic of equal units of measurement. This is called the "interval" scale. Its basic requirement is the empirical determination of equality of intervals, making it possible not only to show differences among objects but also to show whether these differences are equal to each other or not. That is, a difference between *A* and *B* may be 5; a difference between *X* and *Y* on the same scale may be 15; thus, the difference between *X* and *Y* is not equal to the difference between *A* and *B*. However, this scale does not permit the interpretation that the difference between *X* and *Y* is three times the difference between *A* and *B*.

The best scale, in addition to combining all the characteristics of the previous ones, starts at an absolute zero point, and is referred to as the "ratio" scale or a type called "fundamental measurement." Only with this scale is it possible to make use of all the mathematical models and statistical processes available for analysis of data. In the various scaling techniques presented in this chapter, an absolute zero point is desirable but not essential, so that either of the latter types of scales is usable.

Even with the best of these scales it is impossible measure the *true* value of anything. In the physical sciences, standard units have been developed, such as the centimeter, the gram, and others, but changes are constantly taking place in their true values due to variations in temperature, pressure, errors of measurement, and other conditions. In the behavioral sciences the identification of standard units has been more difficult and subject to more conditions likely to cause variations in them. Thus, it has been necessary to develop psychophysical methods for determining scale values that are scientifically usable. The more common of these methods of scaling have been identified as the "equal appearing intervals," "paired comparisons," and

"rank order" procedures. Other methods include those of "internal consistency" and "scalogram analysis."

Method of Equal-Appearing Intervals

This method as applied to the computation of scale values is often referred to as the Thurstone technique, since it was L. L. Thurstone who applied the psychophysical method of equal sensed distances, developed over a hundred years ago, to attitude measurement. The term "equal-appearing" intervals may be explained by a classical example taken from experimental psychology, called the Sanford's Lifted-Weights Experiment. A set of Sanford's weights were made up of heavy Manila paper envelopes in which were enclosed pieces of cardboard, thin sheets of lead, and small pieces of paper so that a set of 108 envelopes of identical size ranged in weight from 5 to 100 grams. A judge was asked to lift these by hand, one by one, or in pairs, and to place them in five groups such that the difference in weight between neighboring piles would seem equal to him. Thus, if the number of weights is increased and the total range of weight is small, the judge will reach a point where the difference between any two neighboring pairs is just noticeable. This "just noticeable difference," or *j.n.d.,* was considered by psychologists as representing the same distance between pairs of weights on the scale, and thus constituted a scale unit.

The same technique was applied to the measurement of attitudes by Thurstone to establish scale units. His procedure was to take a large number of statements (130) made by various people about the church. A number of judges were asked to sort these statements into eleven piles such that one extreme pile represented the highest appreciation of the church, the other extreme pile represented the strongest depreciation, and the intervening piles represented an ordering of varying degrees of attitudes. After each judge completed his sorting, the items in the first pile were given the value of 1, the second a value of 2, and so on. When all judges had completed their sorting, tabulations were made of the ratings on each item and median and Q values were computed. The scale value for each item,

determined by the equal-appearing intervals technique, is the *median* category into which it was placed by the judges. The Q value is used as a measure of ambiguity of the item. For a more extended discussion of this technique, the reader is referred to a textbook on psychometric methods by Guilford.[8]

By using these values, 45 items were selected from the total of 130 so as to secure the best possible representation of the full continuum of items possessing the lowest possible Q values. These items were then assembled into an instrument to measure an individual's attitude toward the church. The measure of this attitude is determined by the average of the scale values on the items endorsed by an individual.

Method of Paired Comparisons

In the method of paired comparisons each subject (attitude, product, or other type of value) is judged as better or worse with every other subject in a group, or on a continuum. This method is very satisfactory for use with a small number of subjects, but as the number increases it becomes more and more tedious. The total number of comparisons to be made is $n(n\text{-}1)/2$, where n is the number of subjects. Thus, if comparisons were to be made among six subjects, the total number of pairs of comparisons would be 15; and if comparisons were to be made with 30 subjects, the total would amount to 435. However, Guilford has suggested a method whereby the number of pairs could be reduced by selecting from all subjects a limited number to become standards for the scale. These should be chosen at approximately equal intervals along the scale and they should be among the least ambiguous of the lot.

An example of the procedure for using the paired comparisons method may be illustrated by a study of interests in sporting events. Assume that it is desired to make a scale of interests, or preferences, for watching five kinds of sports events: baseball, basketball, football, hockey, and tennis. These five sports will require a total of ten comparisons. A list of the ten pairs could

[8] J. P. Guilford, *Psychometric Methods*. New York: McGraw-Hill, 1936, pp. 143–163.

be made up in a prearranged scheme, so that no sport is repeated twice in succession, and that every sport appears an equal number of times on the right and left, somewhat as follows:

baseball-basketball	tennis-baseball
football-hockey	hockey-basketball
basketball-football	tennis-basketball
hockey-tennis	football-baseball
baseball-hockey	tennis-football

A representative sample of persons could be selected, each of whom would be asked to check the sport they liked the best in each pair.

In computing the scale values, the first task is to find the proportions of choices each sport received as compared with every other one. A two-way table is made with the names of the sports written along the top, as column headings, and along the left side, as row headings. The proportions are then entered into the cells of the table. The proportion for each sport as compared with itself is assumed to be .500, as it would have been if the members of the sample had been asked to make this judgment. Each column is totaled and the means of the proportions are computed. These means could constitute scale values in themselves, but the corresponding deviates of these proportions as areas under the normal curve provide more refined scale values. The reader is referred to Guilford for more specific details of the computation of these scale values.[9]

Rank Order Method

The rank order method has been one of the most popular and most practical of all the psychometric methods because of the ease with which a relatively large number of subjects can be judged with reference to one another and because of its wide applicability. In this method the subjects are placed in serial order according to the judge's estimation of their merit. The mean (or median) rank for each subject is computed for all judges and yields a pooled final rank order for each of the sub-

[9] *Ibid.*, pp. 217–241.

jects. When the distribution of subjects can be assumed as approximately rectangular in type these pooled rank orders may be used as scale values. But, in many cases, the subjects fall into a bell-shaped or approximately normal distribution which requires a transformation of the ranks to linear scale units. Only a relatively simple method for determining scale values for normally distributed subjects is presented here. The reader is referred to Guilford[10] for other methods requiring more complex statistical treatment.

If the basic assumption is made that each subject is judged in comparison with the group as a whole, linear scale values can be determined from the proportions of judgments given to every subject as compared with the composite standard. The computation of scale values by this method may be illustrated by the table of ranks given to ten pupils by six judges (see p. 223).

Pupil 2 was assigned a rank of 8 by the first judge, which means that he secured preference over 7 other pupils which fell in ranks lower than 8. If he is considered as having been compared with himself, also, he deserves an additional one-half choice, or a total of 7.5. Thus, any assigned rank (r_i) would mean a total of $(r_i - .5)$ choices for the pupil so ranked. When there are several judges (N), the total number of choices becomes $\Sigma(r_i - .5)$ or $\Sigma r_i - .5N$. The probability that he (R_k) will be chosen as greater than the composite standard (CS) of all the pupils is the total number of choices he receives divided by the total number of comparisons made for the entire group, or the number of judges (N) multiplied by the number of pupils (n). This probability may be expressed by the following formula:

$$P_{R_k} > CS = \frac{\Sigma r_i - .5N}{nN}$$

The first three lines below the ranks for the pupils shown in the table give the sum of the assigned ranks (Σr_i), the total number of choices each pupil was superior to others plus an additional half-choice for himself $(\Sigma r_i - .5N)$, and the probability that he will be chosen as greater than the composite standard $(P_{R_k} > CS)$.

[10] *Ibid.*, pp. 244–260.

	Pupils									
Judges	1	2	3	4	5	6	7	8	9	10
1	10	8	9	7	5	6	4	1	3	2
2	8	9	10	7	6	5	3	4	2	1
3	9	10	8	6	7	5	3	2	4	1
4	10	9	8	7	6	5	4	3	2	1
5	10	7	9	8	5	6	2	4	1	3
6	9	10	7	6	8	1	5	4	3	2
Σr_i	56	53	51	41	37	28	21	18	15	10
$\Sigma r_i - .5N$	53	50	48	38	34	25	18	15	12	7
$P_{Rk} > CS$	.88	.83	.80	.63	.57	.42	.30	.25	.20	.117
x/σ	1.18	.95	.84	.33	.18	—.20	—.52	—.67	—.84	—1.19
SV	2.37	2.14	2.03	1.52	1.37	.99	.67	.52	.35	.00

The next step is to find the deviates corresponding to these proportions from a table of deviates for areas under the normal curve. These deviates are given in terms of sigma units from the mean of a normal distribution, or x/σ. The deviate values may be used as scale units, but since they include both positive and negative numbers it it better practice to make all the scale values positive. This is done by placing the zero point arbitrarily at the lowest scale value, or in this case by adding 1.19 to each of the x/σ values. By use of the deviants from the mean of the normal distribution, several other types of scale values may be computed, such as the *T*-scores and stanines referred to in Appendix B.

Other Techniques

The methods of internal consistency and scalogram analysis are applied when each item in an attitude scale, for example, has become a scale in itself and the total scale is actually a battery of item scales. Items from a preliminary form of the scale, as illustrated by the Thurstone technique, are selected on the basis of how they will relate to the score on the composite of items or the scale as a whole. Scalogram analysis is based upon the types of response patterns shown graphically by the responses to individual items when they are arranged in order of the total score for all items. These techniques are discussed in detail, with illustrative examples, by Goode and Hatt, Guttman, and Likert in the references listed at the end of the chapter.

SELECTED REFERENCES

Bogardus, Emory S. "Social Distance and Its Practical Implications." *Sociology and Social Research* (1933), *17*:265–271.

California State Department of Education. *Evaluating Pupil Progress* (1952), Bulletin, Vol. XXI, No. 6. Sacramento, Calif.: The Department, 1952. (Read pp. 97–124 for rating scales, and pp. 145–151 for sociograms.)

Edwards, Allen L., and Kilpatric, Franklin P. "A Technique for the Construction of Attitude Scales," *Journal of Applied Psychology* (1948), *32*:374–384.

Festinger, Leon. "The Treatment of Qualitative Data by Scale Analysis," *Psychological Bulletin* (1947), *44*:149–161.

Forsythe, Elaine, and Katz, Leo, "A Matrix Approach to the Analysis of Sociometric Data," *Sociometry*, *9*(4):340–47, November, 1946. (The development of the sociometric matrix and sociograph.)

Goode, William J., and Hatt, Paul K. *Methods in Social Research.* New York: McGraw-Hill, 1952. xiv + 386 pp. (Read pp. 232–295.)

Goodenough, Ward H. "A Technique for Scale Analysis," *Educational and Psychological Measurement* (1944), *4*:179–190.

Gronlund, Norman E. *Sociometry in the Classroom.* Harper & Row, 1959. 340 pp. (A comprehensive integration and interpretation of sociometric research literature, and lucid presentation of the principles and procedures of sociometry that have implications for educational practice.)

Guilford, J. P. *Fundamental Statistics in Psychology and Education.* New York: McGraw-Hill, 1950. xiii + 633 pp. (Read pp. 544–576.)

Guttman, Louis. "The Cornell Technique for Scale and Intensity Analysis," *Educational and Psychological Measurement* (1947), *7*:247–280.

Guttman, Louis, and Suchman, Edward A. "Intensity and Zero Point for Attitude Analysis," *American Psychological Measurement* (1947), *12*:57–67.

Jennings, Helen M. *Sociometry in Group Relations: A Work Guide for Teachers.* Washington, D.C.: American Council on Education, 1948. ix + 85 pp.

Loomis, Charles B., and Pepinsky, Harold B. *Sociometry: Theory and Methods.* Sociometry Monographs, No. 20. New York: Beacon House, 1949. 27 pp.

Likert, Rensis. "A Technique for the Measurement of Attitudes," *Archives of Psychology* (1932), No. 140. (Rationale for the development of rating scales with discrete units.)

Lundberg, George A. "The Measurement of Socio-Economic Status," *American Sociological Review* (1940), *5*:29–39.

McNemar, Quinn. "Opinion-Attitude Methodology," *Psychological Bulletin* (1946), *43*:289–374.

Messick, Samuel J. "Some Recent Theoretical Developments in Multidimensional Scaling," *Educational and Psychological Measurements*, *16*:82–100, Spring, 1956.

Moreno, J. L. (Ed.). *The Sociometry Reader*. Glencoe, Ill.: The Free Press, 1960. 771 pp. (A comprehensive treatise on sociometry.)

Park, Robert E. "The Concept of Social Distance," *Journal of Applied Sociology* (1902), *8:*339–344.

Ryan, Thomas. "Merit Ratings Criticised," *Personnel Journal* (May, 1945), *24:*6–16. (A criticism of graphic rating scales.)

Sanford, Fillmore H. "The Use of a Projective Device in Attitude Surveying." *Public Opinion Quarterly* (Winter, 1950–1951), *14:*697–709. (Discusses the use of "cartoonlike" procedures for determining attitudes.)

Selltiz, Claire; Jahoda, Marie; Deutsch, Morton; and Cook, Stuart W. *Research Methods in Social Relations*. Rev. ed. New York: Holt, Rinehart and Winston, 1959. 622 pp. (See index.)

Siegel, Signey, and Shephard, Irma L. "An Ordered Metric Measure of Social Distance," *Sociometry,* December, 1959, pp. 336–343.

Thomas, R. Murray. *Judging Student Progress*. New York: Longmans (David McKay & Co.), 1954. xii + 421 pp. (Read pp. 200–224.)

Thurstone, L. L., and Chave, E. J. *The Measurement of Attitudes*. Chicago: University of Chicago Press, 1929. (Read all.)

Young, L. L. "Sociometric and Related Techniques for Appraising Social Status in an Elementary School," *Sociometry* (1947), *10:*168–177.

CHAPTER X

Use of Computing and Data-Processing Machines in Analyzing Data

Varieties of Punch Cards

Edge-Punched Card System

The Punch Card System

Materials, equipment, and operations
Coding the data for processing
Transmutation of codes
Availability of specialized statistical services

For many years the punched-card system, also variously known as *Hollerith* (developed by Dr. Hollerith around 1885), *Tabulating Machine,* or *Electronic Accounting Machine* system, has made it possible for researchers to expand the scope of their investigations and has radically altered the techniques of scientific inquiry. Statistical projects, formerly prohibitive in time and money, have now become practical.

No recent technological developments for the processing of scientific research data are more promising than those involving machines capable of electronic processing. Already a series of electronic machines have been developed and produced which have made significant contributions to government, science, business, and industry.

The first large-scale digital calculator, the *Automatic Sequence Controlled Calculator,* has been operated successfully at Harvard University since 1944. This machine is credited with starting intensive developments in the field of the "giant brains." In 1948 the first large electronic digital calculator, the *Selective Sequence Electronic Calculator,* was installed by the International Business Machines Corporation in New York City. In 1953 a new series of machines, the *Electronic Data-Processing Machines,* 25 to 30 times more powerful in speed and capacity than the previous ones, were introduced and put on a production-line basis. These machines include the *Electronic Analytical Control Unit,* which performs the computing and control functions by means of electronic pulses emitted at speeds up to 1-millionth of a second; an *Electrostatic Storage Unit* (memory unit), having a capacity of over 10,000 digits which may be stored or obtained from it in 12-millionths of a second; a *Magnetic Drum Storage Unit* (memory unit), having a capacity of over 80,000 digits which may be stored or obtained in 40-thousandths of a second; and a *Magnetic Tape Reader and Recorder* (memory unit), which can introduce and record over 12,000 digits a second. The successful application of these fast, versatile machines is dependent, however, upon the investigator's ability to predetermine in complete detail the procedures to be used in solving the problems to which the machines will be applied.

Recently data-processing centers[1] have been established which provide computer facilities, a processing system, and the necessary services for efficient computation of data-processing problems associated with academic research projects. They also provide assistance in writing computer "programs" which are compatible with their resources and the needs of the investigator.

Because automatic digital computers operate much faster than hand computers and have different logical characteristics, they have transformed numerical analysis. The greater speed

[1] For example, the Western Data Processing center established in 1956 at the University of California at Los Angeles.

and capacity of these machines have made possible the use of techniques which were not feasible previously and have stimulated the search for new algorithms as more difficult problems are posed for solution. Since the information-handling aspect of computers is becoming more and more widespread, we no longer speak of computers, but of *data-processing systems*. The researcher should become familiar with some of the recent literature on computers and data-processing systems; the variety of this literature can be attributed to the manifold applications of these instruments in our everyday life.[2]

The scope of this text does not permit an elaborate presentation of the processes used in these large-scale data-processing systems. The researcher should consult technical specialists for these purposes. However, since nearly all processes make use of the punch card as a basic tool, we present here a brief discussion of the varieties of punch cards, the equipment readily available for most student research projects, and some of the techniques which enable efficient analyses of data by manually operated and electrically driven mechanical devices.

VARIETIES OF PUNCH CARDS

The simplest system for sorting and analyzing data, other than the hand tabulation of schedules and questionnaires, is one which involves the use of the *edge-punched card*. An outgrowth of business accounting, the edge-punched card was invented by Charles F. McBee in the early 1930s and is marketed under the name "Keysort."[3] A similar card has been developed and is marketed under the name *Hadley* cards.[4]

[2] Franz L. Alt (Ed.) *Advances in Computers*, Vol. 1. New York: Academic Press, 1960. 316 pp. Andrew D. Booth (Ed.) *Progress in Automation*, Vol. 1. New York: Academic Press, 1960. 231 pp. Robert H. Gregory and Richard L. Van Horn, *Automatic Data-Processing Systems*. San Francisco: Wadsworth, 1960. 705 pp. Anthony Ralston and Herbert S. Wilf (Eds.) *Mathematical Methods for Digital Computers*. New York: John Wiley, 1960. 293 pp. George P. Shultz and Thomas L. Whisler (Eds.) *Management Organization and the Computer*. Glencoe, Ill.: Free Press, 1960. 257 pp.

[3] Royal McBee Corporation, Westchester Avenue, Port Chester, New York.

[4] Burroughs Corporation, The Todd Company Division, Rochester, New York.

A more elaborate system, and one amenable to electronic data processing, is that based on the punch cards and equipment developed by the International Business Machines Corporation (IBM),[5] Remington Rand,[6] and the Samas Punched Card Division of the Underwood Corporation.[7]

EDGE-PUNCHED CARD SYSTEM

The edge-punched card system of locating data according to predetermined classifications makes it possible for a researcher to compile data from a number of cases in a sample without the use of expensive equipment. This system is satisfactory for classification, cross classification, and sorting of data for small samples or for large samples in which the amount of data is not extensive. The edge-punched card may carry the basic information needed on the card itself, and a respondent may mark his responses to a questionnaire directly on the card. Thus, the information can be read without coding it according to any numerical system.

Cards are available in many sizes, varying from 2 x 3½ inches to 8½ x 11 inches, to fit the needs of various research studies. Figures 23 to 26 illustrate some of the types of cards and the purposes to which they have been adapted. Figure 23 shows a card that has been printed with the basic information for follow-up activities of a commercial enterprise. It provides space for the name of a customer, or prospective customer, his address, phone number, dates for contacting him, and other basic information. Figure 24 shows an adaptation of a card for use in a medical doctor's office. Each bit of information on the face of this card is numbered. The holes punched in the edges of the card are also numbered to correspond with the items printed on the card. At the time a patient is examined by a doctor the card may be filled out by pen, pencil, typewriter, or accounting machine. The holes corresponding to the data recorded are notched with either a hand notcher or electric keyboard notcher, and the

[5] International Business Machines Corporation, 590 Madison Ave., New York, N.Y.

[6] Remington Rand, 315 4th Ave., New York, N.Y.

[7] Underwood Corporation, 1 Park Ave., New York, N.Y.

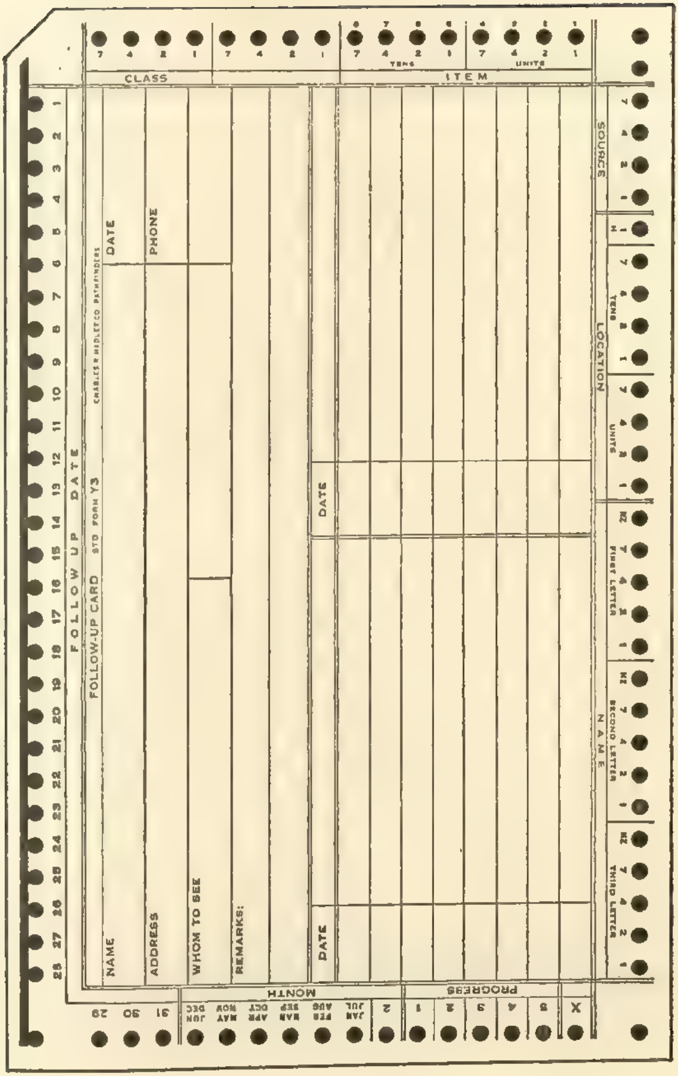


FIGURE 23. Follow-Up Card.

COL. W | SEX M | 1-3 | 4-9 | 10-16 | 17 21 | AGE GROUP

YEAR ENTERED: TENS 7 4 2 1 | UNITS 7 4 2 1

CASE NUMBER: TEN THOUSANDS 4 2 1 | THOUSANDS 7 4 2 1 | HUNDREDS 7 4 2 1 | TENS 7 4 2 1 | UNITS 7 4 2 1

MAJOR CLASSIFICATION

NAME

BIRTH DATE

CLINIC NO.

ADDRESS

COUNTY

PHONE

SEX

RACE

REFERRED BY

DOCTOR

DATE FIRST EXAMINED

A - CONGENITAL ANOMALIES (All Hereditary Defects)	E - CEREBRAL PALSY	I - POLIOMYELITIS
1 - Bones	35 - Pyramidal	69 - Acute
2 - Muscle	36 - Atheto d	70 - Chronic
3 - Nerve	37 - Ataxic	71 - Encephalitic Type
4 - Central Nervous System	38 - Mixed	72 - Spinal Type
5 - Mixed	39 - Rigidities and Tremors	73 - Bulbar Type
6	40 - Congenital Anomalies	74 - Mixed Type
7 - Club Foot	- and Hydrocephalics	75 - Deformities.
B - DEVELOPMENT ANOMALIES	41 - Treated - Outpat. Dept.	76
8 - Variation From Normal	42 " - Nursery School	77
9 - Periadic Departure	43 " - Hospital School	78
10	44	**J - SCOLIOSIS**
11	45	79 - Idiopathic
	46	80 - Congenital -

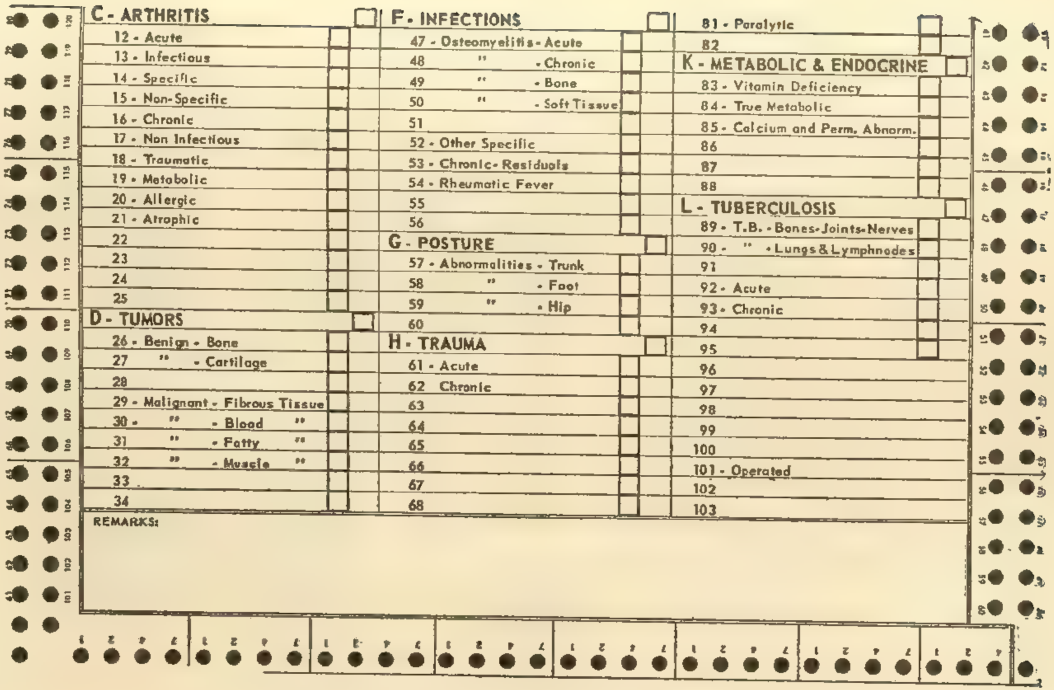


FIGURE 24. Medical Diagnosis Card (8½ x 11″).

card is filed with cards for other patients. Whenever this card is wanted, or the cards for all patients having a specific diagnosis are wanted, it is necessary only to insert a "needle" through the hole numbered according to the desired information and to lift up the pack of cards. The notches in the cards permit the *wanted* cards to "drop out" for analysis. This procedure solves many problems without costly machine investment. Its main advantage is that it permits certain cards with particular classifications to be isolated from a larger number of cards having other classifications. The flow of information into the cards and the procedure for obtaining information from the cards is illustrated by Figure 25, showing the *Hadley* cards. Blank cards, which may be typed, mimeographed, or printed with the specific information desired by a researcher, may be obtained in many sizes for a few cents per card.

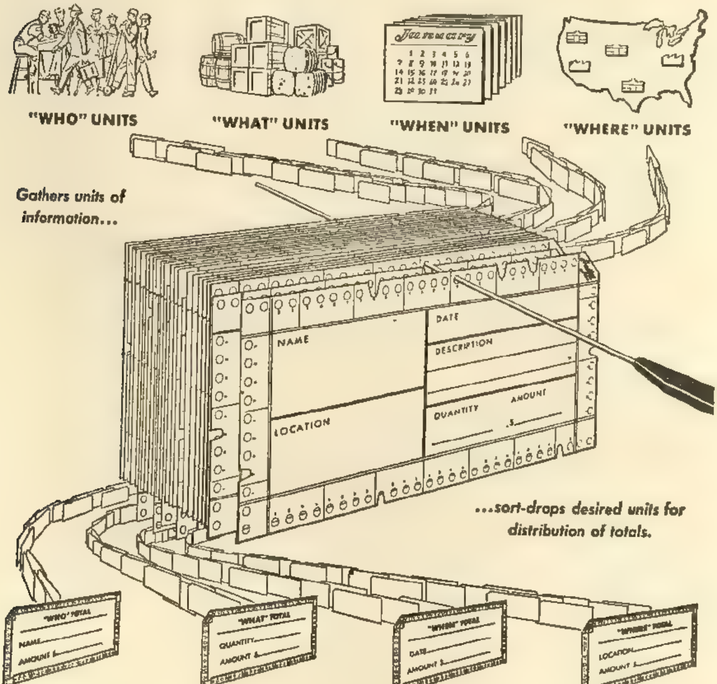


FIGURE 25. Hadley Edge-Punched Cards.

Assume that a researcher wishes to identify the variables affecting retail florists' sales in a given geographic region. He may desire to test a hypothesis that sales increase more rapidly than the rise in per capita income but that this differential varies directly with the size of cities. The data needed to test this hypothesis would consist of a series of yearly reports from florists and estimates of per capita income for each year in the several cities of different sizes. Once he had acquired his data, the investigator could place the pertinent information about each florist on a single card. Some of the necessary information could be written directly on the cards and other information could be coded from a list of information. For example, the size of the different cities could be coded as follows:

Size of City	Hole Number
Under 1,000	1
1,000–4,999	2
5,000–9,999	3
10,000–24,999	4
25,000–99,999	5
Over 100,000	6

All cities ranging in population from 10,000 up to 25,000 would be coded as number 4, and hole number 4, on the top right of the card, would be notched, as illustrated in Figure 26. When all cards representing the sample have been coded and notched, it is possible to "drop out" the cards according to the desired range of population. By successive sorting of the cards dropped out according to other categories, all cards classified according to selected categories may be isolated for analysis. Also this system eliminates the need to replace the cards in the pack in any prescribed or sequential order.

Another important advantage of this system is that all the data for a project may be classified and recorded permanently on the cards. If further analyses, using more complex operations, are desired later the information is readily available.

The edge-punched card system operates on both number codes and letter or word designations. If it is necessary to

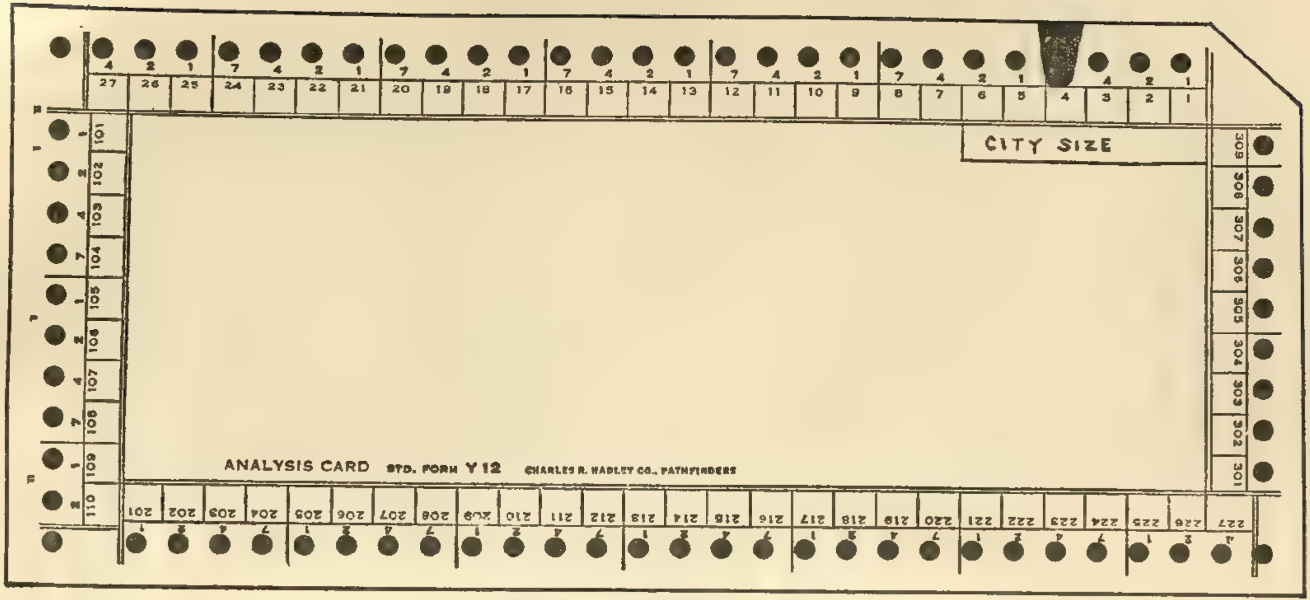


FIGURE 26. Blank Edge-Punched Card.

record numerical values, they may be indicated on the card by using the row marked with boxes containing four holes numbered 7, 4, 2, and 1. Each box can represent one digit from zero to 9 by making one or two notches as follows:

Desired Digit	Notched Holes
1	1
2	2
3	1 and 2
4	4
5	4 and 1
6	4 and 2
7	7
8	7 and 1
9	7 and 2

Larger numbers can be indicated by using adjacent boxes for tens, hundreds, thousands, etc., digits. Any zero value would be indicated by leaving the corresponding box blank.

Cards can be placed in a numerical order by needling each hole successively from *right to left,* starting with the number 1 hole in the units box or field. As each hole is needled the cards that drop out are placed *behind* the cards that remain on the needle. This system also insures that all cards numbered alike (duplicate cards) are placed together. Further details about the edge-punched card system may be obtained from the companies producing the cards. These companies have offices in the principal cities of the United States and in many foreign countries.

THE PUNCH CARD SYSTEM

The *punch card system,* as distinguished here from the *edge-punched card system,* makes use of electrically driven machines developed by such companies as the International Business Machines Corporation, Remington-Rand, and Underwood.

Materials, Equipment, and Operations

Tabulating Card. The basis of the punch card system is the tabulating card. These cards are made of high quality paper

having high tensile strength and a smooth surface of uniform thickness which does not conduct electricity. They are printed with special cylindrical electrotypes to assure perfect registration and legibility. The cards are 3½ by 7⅜ inches in size,[8] as shown in Figure 27. All pertinent information needed in the analysis of a research study is transcribed from the source records—interview schedules, questionnaires, or records of observations—to the tabulating cards in the form of punched holes. The holes are punched in predetermined positions on the cards according to the information registered and serve to actuate the various machines into which they are placed for computational, sorting, or tabulating procedures.

There are 80 columns of digits across the length of the card. Each column contains twelve punching positions, two at the top of the card not indicated by numbers and ten below indicated by the printed digits 0 to 9, to correspond with the numerical data to be punched. Thus, the card contains 960 positions for recording of numerical data or data that may be numerically coded. If a date is to be recorded on a card, for example, it would require six columns. That is, a date such as December 25, 1962 would be indicated as 12-25-62 by punching the corresponding digits in six adjacent columns.

Punching Machines. Tabulating cards are perforated by means of a punching machine actuated by a keyboard (numerical or alphabetical) corresponding to the possible positions on a card. As a key is depressed a hole is cut in one column, and the card is advanced automatically to the next column to be punched. Punching machines are available for either hand or electric operation. Models are available with duplicating features which enable a machine to be used as an automatic coding device by use of precoded master cards. Other models and attachments include the following:

[8] The card illustrated is produced by the International Business Machines Corporation. Remington-Rand produces a card of the same size with 90 columns, using round holes and equipment similar to that used by IBM. The Samas Division of Underwood Corporation produces a smaller card, 2 by 4¾ inches, using round holes, containing 40 columns, and equipment similar to that made by the other companies. (The illustrations of "IBM" cards are reproduced through the courtesy of the International Business Machines Corporation.)

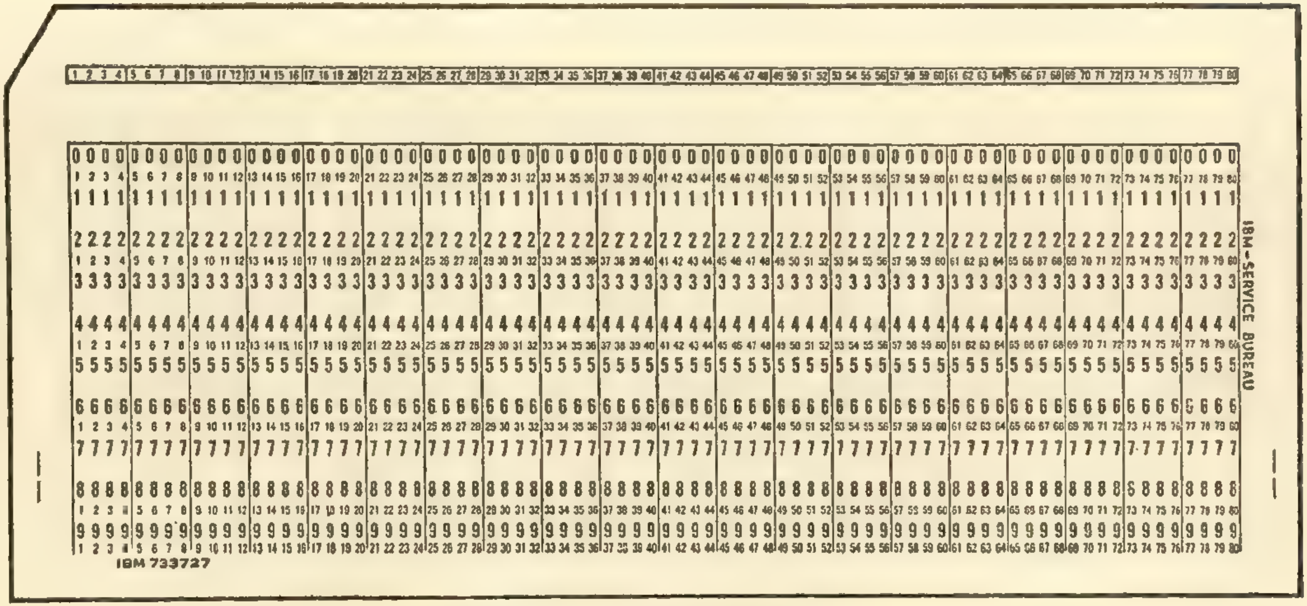


FIGURE 27. IBM Tabulating Card.

Printing card punch—equipped with printing mechanisms which automatically interpret the punched information at the top of the card directly above the hole being punched.

Duplicating punch—for automatic punching of repetitive information from a master card into a group of succeeding detail cards.

Gang punching—for automatic copying of punched information from one card into one or more detail cards following it, that is, the punching of a large number of cards in identical manner.

Reproducing punch—for copying all or part of the information from one card to another. Information from one set of punched source cards is automatically punched into another set of cards by having two decks of cards fed through the machine synchronously.

Mark-sensed punching—for automatic punching of a card by means of electrically-conductive marks made on the card with a special pencil. The original information is marked on the card with an electrographic pencil and this information is then translated directly into punched-hole form.

Summary punching—for the automatic conversion into punched-hole form of information developed by the tabulating and accounting machine described later.

Calculating punch—for computing a result by multiplication, division, addition, or subtraction. Any combination of these operations can be performed with automatic checking to prove the accuracy of the calculations.

Verifying Machine. Cards punched may be checked for accuracy by use of the verifying machine. A second operator verifies the original punching by depressing the keys of a verifier while reading from the same source data. The machine compares the key depressed with the hole already punched in the card. A difference causes the machine to stop, indicating a discrepancy between the two operations. A notch in the upper right edge of the card indicates that it has been key-punched and verified. A notch directly above a column signifies that the punching of that column is in error. By use of this verifying process a researcher can be assured of almost complete accuracy in the transcribing of information from his original records to cards.

Automatic Alphabetic and Numerical Interpreter. The in-

terpreter machine makes it possible to read easily the numerals represented by the punched holes. This is helpful in checking, filing, selection, and reference operations. A model is available that is a combination check writer, check protector, and interpreter. The translation of information can be made by use of the interpreter machine at a rate varying from 60 to 100 cards per minute.

Sorting and Collating Machines. The punch cards may be punched in any order, but analysis of them requires that they be arranged in sequence according to some desired classification. A total pack of cards may be placed in the *sorting machine* and sorted into a number of different categories. As the cards pass through the machine, contacts are made through the holes which cause the cards to be dropped into pockets corresponding to the numbers required by the predetermined classifications. The automatic sort is made on one column at a time, but since the cards pass through the machine at rates varying from 250 to 1000 cards per minute, sorting according to any classification requires but a very short time. Counting devices are available for these machines so that the number of cards may be registered for each of the 12 punching positions, for the subtotal of all cards on one sort, and for the grand total of cards on several sorts.

The *collating machine* makes it possible to interfile cards from different files or to match groups of cards in one file with groups in another file. It may be used, also, for selection of cards of a given classification or range of classifications.

Tabulating and Accounting Machines. An electric tabulating and accounting machine performs the operations of adding, subtracting, and printing. As the cards pass through the machine at a rate up to 150 cards per minute, a number of columns of data may be subjected to multiple addition, subtraction, classification, and printing at one time. One or more copies of detailed listing of data or summary tables can be produced in a single run of the tabulation cards. A summary punching machine may be attached electrically to the accounting machine so that as the detailed printing of cards is made on the account-

ing machine subtotals and totals for various classifications may be transferred to the summary punch for punching of cards carrying the summary data for each classification.

Electronic Statistical Machine. Statistical work is essentially a process of counting units in many different classifications. It is frequently desirable, at the same time this counting is being done, to obtain subtotals for various classifications, check for accuracy or consistency, and balance counts to the major totals to check for accuracy of summaries. All these functions are performed by the electronic statistical machine when producing printed summaries on reports.

Card-Programmed Calculating Processes. By connecting several machine units, several operations may be handled simultaneously. This process requires a set of instructions to activate different operations of the machines. The *accounting machine* reads from the punched cards the factors of data for calculation and the codes that instruct the machines about calculations to be made. The factors are introduced into counters of the several machines, and calculations are made according to coded instructions on a program card. A *storage unit* makes possible the holding of figures until they are needed in calculations. Upon completion of the calculations, results may be printed on a report by the accounting machine, as well as punched into a card by the punch unit of the calculating punch.

Typewriter Tape Punching and Reading. Typewriter tape punching is a means of recording information in code upon a tape by use of a special typewriter. As a document is being created on a typewriter, any or all of the typed information can be recorded on a multiple-channel tape. *Tape reading* is a process of feeding coded tapes through a tape-to-card punch to convert the coded information into punched cards.

Data Processing. Data processing, from a machine standpoint, entails entering a complete set of instructions as well as the initial source data into a series of machines to enable them to arrive at the completed final results or reports in one operation. This type of data processing requires the programming of each step in the procedure—including the solution to all excep-

tions—before source data are processed. Through the use of punch cards, magnetic tapes, magnetic drums, electrostatic storage, and printing units, the machines are capable of high-speed production and processing of large procedures and complex problems.

Coding the Data for Processing

In preparing a survey instrument or schedule for analysis by the punch card system, it is desirable to determine the code to be used for transferring each item of information to the tabulating card prior to printing the survey instrument. The punch card, itself, is the end product of such coding.

The questionnaire, illustrated in Figure 28, entitled *Your Future Education,* shows the code numbers above the response positions. While it would have been possible to allow spaces on the tabulating card for each respondent's name, the names themselves were of no consequence and only the questionnaire number was used. The number occupying the first four columns of the tabulating card was 2899 for the questionnaire illustrated. This instrument was four pages in length, with the pages offset so all responses could be seen by the punching machine operator without turning the pages of the questionnaire. Column 5 on the tabulating card was punched with digit 1, 2, 3, or 4 to indicate the quarter of the year in which the questionnaire was answered. Number 6 was punched with digit 1 or 2, for male or female, respectively. Two columns, 7 and 8, were allowed for a meaningful code number representing the school in which the respondent was enrolled. Responses could be recorded for as many as 100 schools, each of which was assigned a number from 00 to 99. Column 9 was used to indicate with a digit from 1 to 5 the response to item I. In columns 10 to 14 only the digits 1 or 2 were used for "yes" or "no" answers to the five questions in item II. The remaining questionnaire items, for which the responses are shown in the three columns to the right, were coded in a like manner. Each item enclosed in brackets, instead of parentheses, required the insertion of a code number (as for name of school) prior to punching the tabulating card.

Nº 2899

Quarter________ (5)

(TO BE FILLED OUT BY PRINCIPAL)

Name________________________Sex: Male__ Female__ (6)

School______________________________ (7-8)
[ ]

YOUR FUTURE EDUCATION

Information on this questionnaire will be kept confidential. By answering the questions carefully and honestly you will be helping yourself and all Oregon high school students.

I

Check (X) *one* of the following:

1. I am *not interested* in entering college next Fall. (9) ( )1

   If you *would like* to enter college next fall check *one* of the following which most nearly applies to you:

2. I am *not sure* I will be able to meet the expenses of going to college next year. ( )2

Page 2

Yes No
( ) (15) ( )
( ) ( )
( ) ( )
( ) (18) ( )

(19-20) [ ]
(21-2) [ ]
(23-4) [ ]

Page 3

(28-9) [ ]

______ yrs.

Yes No
( ) 32 ( )
( ) ( )
( ) ( )
( ) (35) ( )

(36-8) [ ]

(39-40) [ ]

(41) [ ]

Page 4

(44)
( )1
( )2
( )3
( )4
( )5
( )6
( )7
( )8
( )9
( )0

(45)
( )
( )
( )

(48)
( )
( )
( )

3. I do *not* think I will be able to meet the expenses of going to college next year. ( )3

4. I know just about how my college expenses will be met for next year. ( )4

5. I will not enter college because of *personal* or *other* reasons. The reason is *not mainly* a lack of money. ( )5

II

Read the following questions carefully and check (X) your answer *Yes* or *No.*

	Yes	No
1. Do you *prefer* to take a job rather than enter college?	( )[10]	( )
2. Are you planning to enter one of the armed services?	( )	( )
3. Do your parents oppose your going to college?	( )	( )
4. Do you expect to be married within the next year or are you married now?	( )	( )
5. Will you have to help support your family next year?	( )[(14)]	( )

(25)
( )1
( )2
( )3
( )4
( )5

(28-7)
_____ yrs.

(42)
( )1
( )2
( )3
( )4
( )5
( )6
( )7

(43)
( )1
( )2
( )3
( )4
( )5
( )6
( )7
( )8
( )9
( )0

(51)
( )
( )
( )

(54)
( )
( )
( )

(57)
( )1
( )2
( )3
( )4
( )5

(58)
( )1
( )2
( )3
( )4
( )5

FIGURE 28. A Coded Questionnaire.

The numbers punched in the tabulating card comprise a thoroughly meaningful code number representing a series of variables. On the basis of these code numbers each person can be differentiated from every other person as each person is represented by a different or unique number. On the basis of the above questionnaire, a meaningful code number of 2899312-6422222 . . . would be interpreted as follows:

2899 One individual, identified by this number from a predetermined coded name list.
3 He answered the questionnaire in the third quarter of the year.
1 Male.
26 He attended a specific school identified by this number from a predetermined coded school list.
4 He knows how his college expenses will be met next year.
2 He does not wish to take a job rather than go to college.
2 He does not plan to enter the armed forces next year.
2 His parents do not oppose his going to college.
2 He does not expect to be married within the next year.
2 He will not have to help support his family.

The remaining 58 items for this questionnaire were coded and could be interpreted accordingly. This example illustrates only one of the numerous ways in which data may be coded for the punch card system of statistical analysis. However, a researcher should *always* consult a technician specialized in this method of analysis for the most efficient system of coding for his particular research study. In some larger universities graduate students may elect or are required to take a course in machine data processing before doing their required research.

A procedure used by the Registrar's office at the University of Oregon, and illustrated by the *IBM Coding Card* and *Student Master Card* in Figures 29 and 30, is based on information obtained from the admissions applications of students at the time of matriculation in college. A clerk prints, by hand, on the coding card the information given by the student on the application blank and any other pertinent information available in the Registrar's office. From the coding card a *Student Master Card* is punched giving of the following information that which is available at the time.

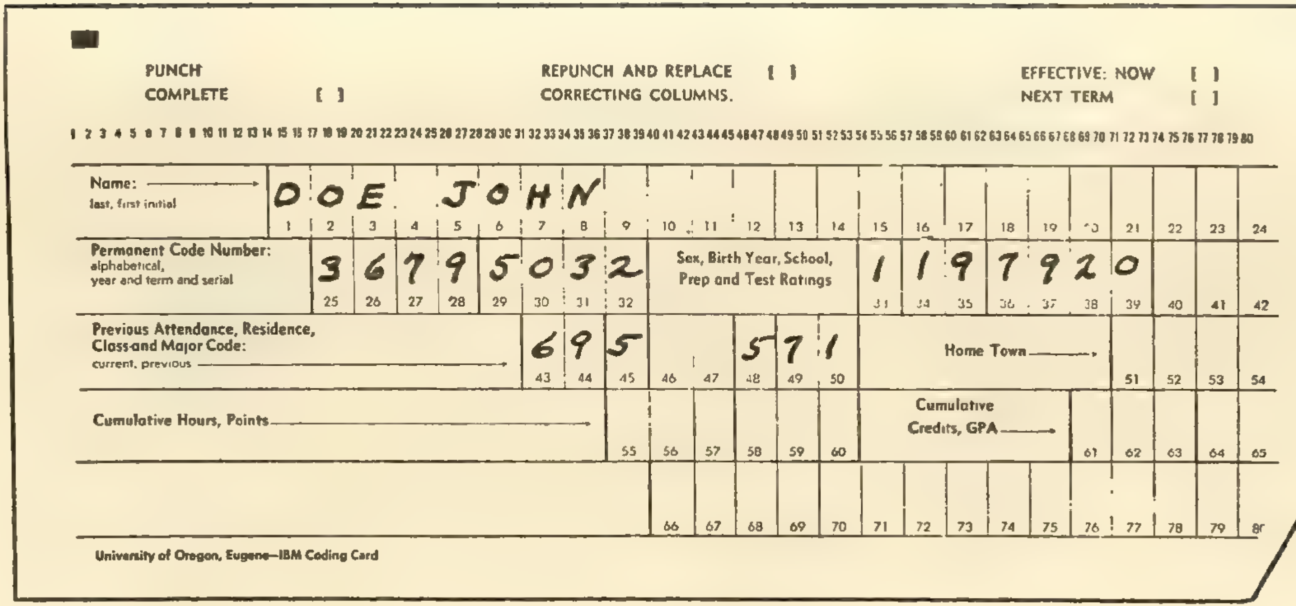

PUNCH
COMPLETE [ ]
REPUNCH AND REPLACE [ ]
CORRECTING COLUMNS.
EFFECTIVE: NOW [ ]
NEXT TERM [ ]
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80
Name: last, first initial — DOE JOHN
Permanent Code Number: alphabetical, year and term and serial — 3 6 7 9 5 0 3 2
Sex, Birth Year, School, Prep and Test Ratings — 1 1 9 7 9 2 0
Previous Attendance, Residence, Class and Major Code: current, previous — 6 9 5 5 7 1
Home Town
Cumulative Hours, Points
Cumulative Credits, GPA
University of Oregon, Eugene—IBM Coding Card

FIGURE 29. Coding Card.

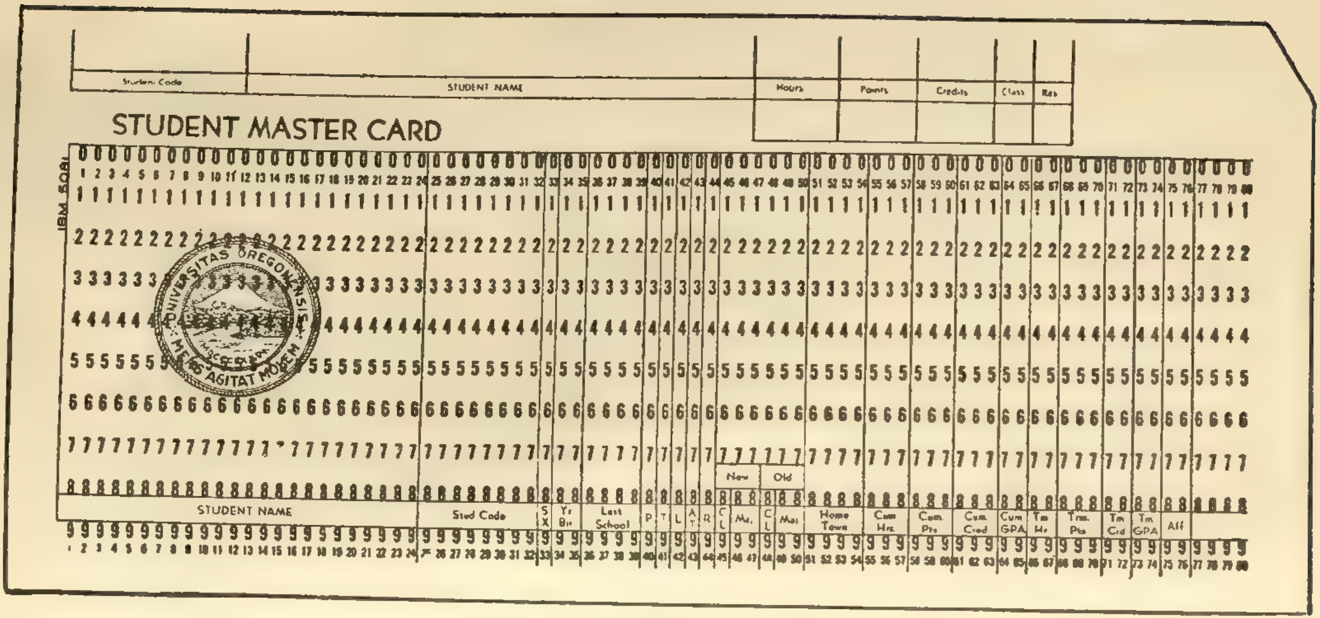
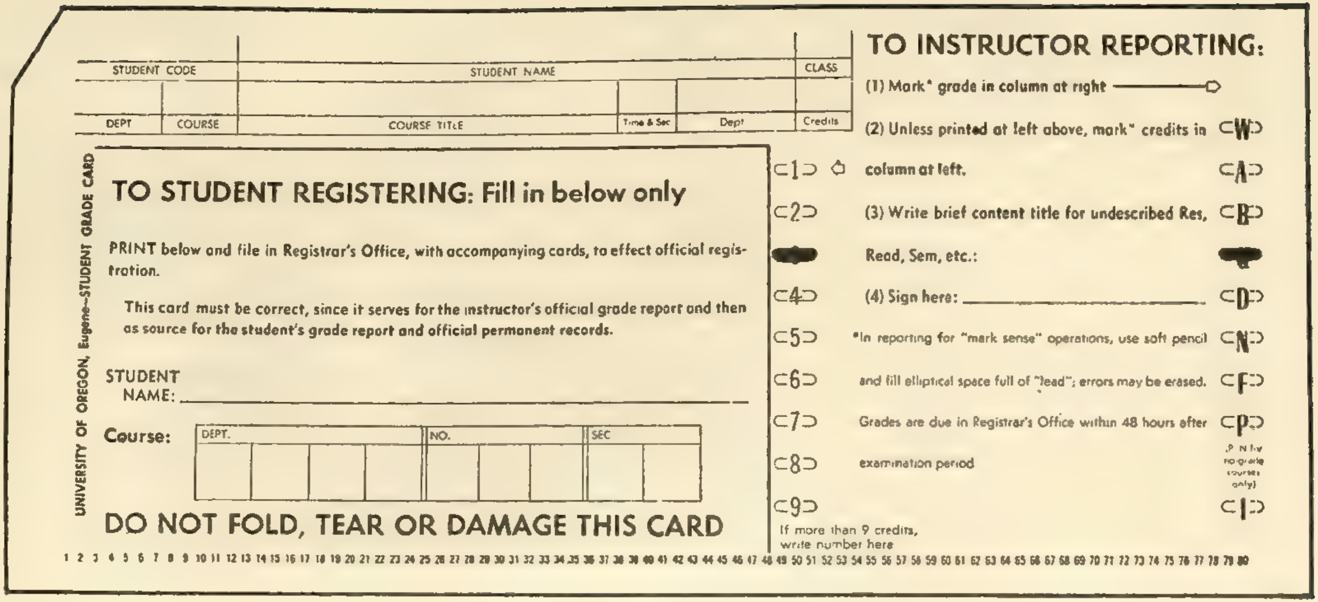


FIGURE 30. Student Master Card.

Columns	Data
1–24	Student's name (numerically coded from a code list).
25–32	Student's college code number (admission number).
33	Sex.
34–35	Year of birth.
36–39	Last school attended (from a coded list of schools).
40	Preparatory school rating (in stanines).
41	College aptitude test rating (in stanines).
42	Language—two years in high school (from a coded list of languages).
43	Year of previous attendance in college.
44	Residence or nonresidence status.
45	Present class: Freshman, Sophomore, Junior, Senior, Graduate.
46–47	Present major study.
48	Previous class: (as column 45).
49–50	Previous major study.
51–54	Student's home town (from a coded list of towns).
55–57	Cumulative credit hours attempted.
58–60	Cumulative grade points earned.
61–63	Cumulative credit hours completed.
64–65	Cumulative grade point average.
66–67	Present term hours enrolled.
68–70	Present term grade points earned.
71–72	Present term hours completed.
73–74	Present term grade point average.
75–76	Fraternity or sorority affiliation (from a coded list).
77–80	Unused columns.

The *Student Master Card* would be repunched when other information, such as term grades, becomes available. The grade reports are made on *Mark-Sensed Cards* which have been previously punched and printed (by use of the reproducing punch and interpreter) from the *Student Master Card* with information for the instructor reporting a grade. The *Grade Card,* illustrated in Figure 31, is sent to an instructor at the end of a term of instruction. He marks the number of credit hours for which a student has been enrolled in a course in the numbered column and the grade earned in the lettered column. This information is automatically punched in the card by a mark-sensed



UNIVERSITY OF OREGON, Eugene—STUDENT GRADE CARD

STUDENT CODE		STUDENT NAME			CLASS
DEPT	COURSE	COURSE TITLE	Time & Sec	Dept	Credits

TO STUDENT REGISTERING: Fill in below only

PRINT below and file in Registrar's Office, with accompanying cards, to effect official registration.

This card must be correct, since it serves for the instructor's official grade report and then as source for the student's grade report and official permanent records.

STUDENT NAME: ______________________

Course:

DEPT.				NO.			SEC	

DO NOT FOLD, TEAR OR DAMAGE THIS CARD

TO INSTRUCTOR REPORTING:

(1) Mark* grade in column at right

(2) Unless printed at left above, mark* credits in column at left.

(3) Write brief content title for undescribed Res, Read, Sem, etc.:

(4) Sign here: ______________________

*In reporting for "mark sense" operations, use soft pencil and fill elliptical space full of "lead"; errors may be erased.

Grades are due in Registrar's Office within 48 hours after examination period

⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃

If more than 9 credits, write number here

⊂W⊃ ⊂A⊃ ⊂B⊃ ⊂C⊃ ⊂D⊃ ⊂N⊃ ⊂F⊃ ⊂P⊃ ⊂I⊃

(P N for no-grade courses only)

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80

FIGURE 31. Student Grade Card.

reproducing punch and then automatically punched in the *Student Master Card* from which printed reports may be made by an accounting and tabulating machine. In this manner it is possible to reproduce and send out end-of-term grade reports in large volume in a very short time. Without the use of punch-card equipment, this process would involve a large clerical staff working a number of days to get out the necessary reports.

Transmutation of Codes

If an extensive research study is being conducted, the raw data frequently cannot be contained in the 80 columns provided on an IBM tabulating card. Many times, however, the raw data may be coded in such a way as to make the 80 columns sufficient. In the same manner as scores are grouped in a frequency distribution to simplify statistical computation, the variables to be included on the tabulating cards can be coded into 9 to 12 classes, or intervals, without appreciable distortion of the facts. This makes it possible to punch into one column on the card data that would otherwise require several columns. For example, a series of test scores ranging from 182 to 242, which would require three columns in their raw state, could be coded for entry into one column as follows.

Transmutation Table

Score Intervals	(X) *Midpoint*	(C) *Coded Scores*
183–187	185	0
188–192	190	1
193–197	195	2
198–202	200	3
203–207	205	4
208–212	210	5
213–217	215	6
218–222	220	7
223–227	225	8
228–232	230	9
233–237	235	10
238–242	240	11

The equation for reconverting these coded scores (C) into raw scores (X) is:

$$X = 5C + 185$$

Thus, to find the raw score for a coded score of 6, for example, the computation would be:

$$X = 5(6) + 185, \text{ or } 215$$

All obtained raw scores of 213 to 217 inclusive would be given the raw score value of 215 and would be given a coded value of 6. The coded scores would yield the same means, standard deviations, correlation coefficients, etc., as the original scores do when grouped in intervals of five raw score units. The coded scores of 10 and 11 would be punched in the two blank positions at the top of the tabulating card. Another transmutation procedure would be to convert all score values into *stanines*[9] which would require only the regular nine positions on the tabulating card and would avoid the additional technical inconveniences in some machine operations arising from the use of the two positions at the top of the card.

It is possible to make many more types of codes by recoding the once-coded scores. This process is referred to as multiple coding. It is based on the concept that any score (C) can be expressed in the *binary number system* as:

$$C = a2^n + b2^{n-1} + c2^{n-2} + d2^{n-3} + \ldots$$

where: a, b, c, etc., are each limited to two values, 0 or 1. In the binary number system, each time the digit 1 is moved one position to the left its value doubles, as shown on the next page.

Since only two digits are required, binary numbers may be represented in electronic equipment by vacuum tubes, being either *on* or *off*. The *on* condition may represent "1," and the *off* may represent "0." For instance, assume four vacuum tubes, designated a, b, c, and d, from left to right. If a, b, and d were *on* and C were *off* this would represent the binary figure 1101

[9] Refer to any standard textbook on measurement or statistical methods for procedures for converting values to stanines.

Decimal Numbers	*Binary Numbers*
0	0
1	1
2	10
3	11
4	100
5	101
6	110
7	111
8	1000
9	1001
10	1010
11	1011
12	1100
13	1101
14	1110
15	1111
16	10000

or, according to the table above, the decimal figure 13. This may also be shown by the previous formula as:

$$C = (1)2^3 + (1)2^2 + (0)2^1 + (1)2^0, \text{ or}$$
$$C = 8 + 4 + 0 + 1 = 13$$

Fortunately, machines are designed to take in decimal numbers, convert them to the binary number system, compute in binary, and then deliver the answers in decimal numbers. This is the basic principle upon which electronic statistical calculators have been developed. By this process, the number of many-categoried variables may be multiplied by several times the number that may be punched on a tabulating card by the raw score method or the simple coding method. For applications of this procedure, also, a researcher should consult a specialized technician.

Availability of Specialized Statistical Services

The actual punching, sorting, tabulating, and accounting of data should be done by experienced clerks or technicians rather

than by the researcher himself, unless he has had considerable instruction and experience in the procedures used in the punch-card system. A few business machine corporations lease the various statistical machines used in this method, but the cost is usually too great for a single research study being conducted by a graduate student. Many colleges and universities, however, have statistical departments which use these machines and which are usually willing to make the statistical analyses necessary for their graduate students for the cost of the tabulating cards and the labor necessary for doing the analysis. In addition to college and university departments, there are well over 100 commercial centers throughout the United States and in the principal cities of the world in which a researcher may have his data analyzed at a reasonable cost.

SELECTED REFERENCES

Baehne, George Walter. *Practical Applications of the Punched Card Method in Colleges and Universities.* New York: Columbia University Press, 1935. 442 pp.

Cooley, William W., and Lohnes, Paul R. *Multivariate Procedures for the Behavioral Sciences.* New York: Wiley, 1962. 232 pp.

Goode, William J., and Hatt, Paul K. *Methods in Social Research.* New York: McGraw-Hill, 1952. 386 pp. (Read pp. 313–325.)

Gregory, Robert H., and Van Horn, Richard L. *Automatic Data-Processing Systems.* San Francisco: Wadsworth, 1960. 705 pp. Describes recent developments in computer programming and the applications of computers to various research designs and analyses.)

IBM General Information Manual: Introduction to IBM Data Processing Systems. New York: International Business Machine Corp., 1960. 95 pp.

Ledley, Robert S. *Programming and Utilizing Digital Computers.* New York: McGraw-Hill, 1962. 568 pp. (Good discussion of machine language, programming, and special data processing techniques.)

McCracken, D. D. *A Guide to Fortran Programming.* New York: Wiley, 1961. 88 pp.

Oakford, Robert V. *Introduction to Electronic Data Processing Equipment.* New York: McGraw-Hill, 1962. 340 pp.

Report to Conference on Data Systems Languages Including Initial Specifications for a Common Business Oriented Language (CoBoL) for Programming Electronic Digital Computers. Washington, D.C.: Government Printing Office, April, 1960. 114 pp. (Department of Defense.)

Schultz, George P., and Whisler, Thomas L. (Eds.) *Management Organization and the Computer.* Glencoe, Ill.: The Free Press, 1960. 257 pp.

Van Ness, Robert G. *Principles of Punched Card Data Processing.* Elmhurst, Ill.: The Business Press, OA Business Publications, Inc., 1962. 263 pp. (This book bridges the gap between manual-to-mechanical-to-automated data processing.)

CHAPTER XI

Writing the Research Report

Style

Major Divisions of a Report
- Introductory section
- Presentation of evidence section
- Summary and conclusions section

Forms Used in Presenting Evidence

Writing, Revising, and Rewriting the Report

Criteria for Judging Research Reports

After the researcher has collected his data and analyzed them, his next step is to present his results in a form that can be used by others. The primary purpose of the presentation is to communicate ideas, but the researcher is often handicapped in preparing his report because he does not know just which ideas will interest his readers. Some readers may be concerned only with his conclusions and be relatively indifferent about how they were obtained, whereas other readers may themselves be researchers and wish to study the techniques used, review the analyses to see if they agree with the researcher's conclusions, and perhaps either repeat the experiment or again search the library sources to see if the conclusions can be validated. Still other readers may be searching for suggestions for additional research in the same general area. Unless the researcher is pre-

paring his report for a specific type of reader, he must assume that he is writing for people with varied interests. Therefore he must give a complete and full description of his problem, his method of attacking it, his analyses, and his conclusions. This will enable readers with specialized interests to concentrate on those aspects of his report which are valuable to them.

Often while writing a detailed report the researcher may discover errors or areas that need further research. A careful step-by-step exposition of a research project will greatly help the researcher to clarify his own thinking. Moreover, if he asks colleagues or advisors to read his drafts, they may make suggestions that will be of great value to him.

STYLE

A good report must combine clear thinking, logical organization, sound interpretation, and an effective style of writing. It is most important that a writer feel he really has something worth saying. If he does, he will probably come forth with a report that meets these criteria, but if he is not convinced himself that he has a worthwhile piece of research to report, he will have a difficult time in his writing.

Not only should the research report be written in acceptable English, it should also be as readable as possible. This means that the presentation should not leave the reader puzzled about the writer's meaning and that the reader's interest should be kept in mind by the writer without in any way sacrificing the honesty of the report. It is not easy to tell the researcher how to present his finding in a scientific, yet interesting, fashion. The inclusion of all details that are germane and the exclusion of those that are not frequently presents the writer with difficult choices; for example, should he include anecdotes which may add color but which are not strictly necessary and may distract the reader from the true purpose of the report which is to give information about the research project.

Clarity and accuracy are the foremost requirements for scientific writing. The author must express himself with sufficient precision so that the reader will not misinterpret what he has

said. He must present his material in such a way that there are no gaps in the flow from one point to another. A research report is no place for poetic license, colloquialism, or barbarism. Words should be used in their generally accepted sense, and sparing use should be made of technical terms familiar only to persons with a considerable background in the field. In certain disciplines commonly used words and phrases have a special technical meaning. Inexperienced writers in these areas frequently do not define these words and phrases, which makes it difficult for a reader with a slightly different background to understand the discussion, even if he is actually reasonably well-informed about the subject. Clarity may call for the repetition of the same word more often than would be acceptable if the discussion were nonscientific. Nevertheless, it is possible for a writer to present accurate information in a clear manner and still give a degree of sprightliness to his report.

A research report should be prepared according to the best composition practices. This topic is discussed at greater length in Appendix A—*A Form Manual for Research Reports.* The writer should be careful to make correct use of such aids to readability as proper paragraphing, the use of topical sentences, illustrations and examples, short sentences, and section headings, Footnote references should be in proper form, and the bibliography should be reasonably complete and in proper form.

A single paragraph should not deal with too many thoughts. If a paragraph is either too short or too long, it detracts from the readability of the report. In general, a paragraph should not be longer than half a typewritten page double spaced. Very short paragraphs should be used rarely, and single-sentence paragraphs should not be used at all in this type of exposition. It is desirable to use a topical sentence at the beginning or end of each paragraph to summarize its content.

The researcher should present his data in a manner that will interest his reader while being sure to include all the evidence upon which he has based his conclusions. In deciding how to present data most effectively, two questions must be carefully

considered: (1) What information is to be included? and (2) In what form is it to be shown? Some data may be presented in the form they were obtained, whereas extensive numerical data should be grouped, classified, or tabulated in some summary form. Sometimes the raw data may be too cumbersome to reproduce in the main part of the report. Sometimes not all the data are relevant—for example an analysis might be based on a relatively long interview schedule in which only a portion of the answers are pertinent. In such a case, a summary may be given in the text with a reference to the raw data which would be found in an appendix. However, all data, especially those that are original with the study, must be made available in some part of the report so the reader can study them and decide whether or not he believes the writer analyzed them correctly.

If the data consist of material collected by other writers, the reader should still have the opportunity of reading and analyzing the relevant materials. Footnote references to the original source are sufficient if the source is readily available in a library. In such cases quotations need not be unduly long. The writer should select a few key sentences for quotation and just paraphrase the rest. If, however, his source is not readily available to the ordinary reader because it is in a private library, has not been published, has not been translated, or only a very few copies are in existence, the researcher should include sufficiently long quotations to avoid the charge that he picked a few sentences out of context in order to strengthen his argument.

If the data are numerical, the researcher may wish to use a considerable number of tables to present them. Most elementary texts dealing with statistical techniques include rules for the preparation of tables, and these should be followed carefully. The writer of a research report should not expect a reader to spend any great amount of time puzzling over his tables. It ought to be possible for a reader to follow the discussion with little or no reference to the tables. Yet to accomplish this objective, the writer of the report ought not to repeat the tables in text form. Rather, he should draw conclusions from the tables and assume that if the reader wants to know the bases

for these conclusions, he can turn to the tables and find out for himself.

MAJOR DIVISIONS OF A REPORT

A research report should be divided into the following sections: (1) an explanation of the reasons for the study, (2) an explanation of the methodology to be used, (3) the presentation of the evidence and its analysis, and (4) the presentation of the conclusions to be drawn from the analyses which is often integrated with a summary of the entire study. The number of chapters devoted to each of these divisions depends in part upon the nature of the project, the interests of the anticipated readers, the style of the writer, and a number of other factors. In any case, the largest number of pages will be devoted to the presentation of the evidence and the accompanying analyses.

Introductory Section

This section ought to include the following, although not necessarily in this order:

1. A statement and description of the general problem.
2. A breakdown of the problem into its constituent elements, major subdivisions, or questions.
3. A description of the scope and limits of the project which is being reported and the exact relationship between the project and the problem.
4. A statement of the relationship between this study and other work in the field.
5. An explanation of the need for this study and its importance.

The general purpose of the study should be described early, in the opening paragraphs if possible. The historical background and the present importance of the problem should be presented so that the reader understands its significance. Once the problem has been throughly explained and analyzed, the writer should explain how he will attack it. This portion of the introductory section should include a description of the following:

1. The sources of the data, and their reliability and adequacy.
2. The techniques which will be used to analyze data and how these analyses will help to solve the problem.
3. The assumptions that will be used in the analysis.
4. The technical terminology used.

All this should be described so explicitly that anyone reading the study could repeat the research for purposes of corroboration or refutation of the findings.

By the time the reader has finished the introductory section, he should be ready to study the evidence that is to be presented. He will know exactly how the data are to be handled and the purpose of the project.

Presentation of Evidence Section

The chapters constituting the body of the report will vary in number and length according to the amount and nature of the evidence to be presented. The data themselves should be described quite fully, they should be analyzed in detail, and all the evidence resulting from the analyses should be presented. These chapters are primarily for the use of the reader who wishes to make a detailed study of the problem, therefore, every bit of relevant evidence should be supported by logical reasoning and empirical facts.

Materials should be organized systematically and presented under appropriate headings. For example, the following order might be adopted:

1. Statement of specific question or hypothesis under study.
2. Presentation of relevant data.
3. Interpretation of data.
4. Conclusions and interpretation for each specific question above.

Sometimes a researcher becomes so carried away with some aspects of the details of his study that he loses sight of his major objective. At other times, he may be inclined to expand the discussion of those topics for which he can obtain information easily and slight those for which he can not readily get material. This is particularly unfortunate if the hard-to-get information

is more relevant to the study than that which is easy to obtain. Many graduate students, in writing theses, hate to leave unused any reference card, although the information on it may have no relevance to the specific problem being investigated. To insure that the presentation is "tight" and orderly, the writer may well ask himself some or all of the following questions:

1. Have I answered adequately all the major questions posed in the introductory section?
2. Have I given too much attention to minor questions at the expense of the major ones?
3. Do my answers to the minor questions help to answer the major ones?
4. Have my own interests or prejudices led me into "detours" or played too important a part in my consideration of the problem?
5. Have I eliminated all unessential materials from my report?
6. Have I introduced a bias either through the omission of evidence or through the process by which I obtained the evidence?
7. Have I erred by including too many examples, quotations, and statistics, or ought I to have still more in order to give the reader a clear understanding of the discussion?
8. Do my statistical tables clearly show what I say they do?
9. Does my report look too much like a mathematical exercise?

Summary and Conclusions Section

The summary and conclusions section should be written so that it can stand by itself—in other words it ought to constitute a synthesis of the preceding chapters. It should include a brief review of the problem dealt with, the sources of the data investigated and the methods of analysis used as well as the conclusions or solutions the researcher has reached on the basis of the evidence reported, the application of these conclusions, and the writer's recommendations. The exact amount of detail needed to support the conclusions or solutions depends somewhat on the nature of the presentation made in the previous chapters or sections. Under no circumstances, however, should new material or additional analyses of data be added at this point.

When he is writing conclusions to a specific aspect of the in-

vestigation, the researcher should review his evidence to be sure that it supports fully every generalization, conclusion, or solution. However, it is neither necessary nor desirable to repeat all the evidence that was presented in earlier chapters. When writing his conclusions, it may be helpful for the researcher to ask himself the following questions:

1. Is the quantity and quality of my evidence an adequate base for the conclusions I have drawn?
2. Are my conclusions too broad, in the light of the supporting evidence? Should they be limited or qualified in some manner?
3. Have I omitted any evidence because it might weaken the conclusions I have drawn?
4. Have I stated that relationships were of the cause and effect type whereas the evidence would indicate that one event just followed another without the first necessarily causing the second?

The writer should always remember that his conclusions must be substantiated by the evidence presented. However, he should not become so enamoured of qualifications and limitations that his report appears to have almost no application to real life when, as a matter of fact, it may have considerable value. The following quotation illustrates qualifications made *ad nauseam*.

A city editor explained what libel was to a new reporter, a cub, and told him how careful he had to be in his news items. The young reporter was sent out to cover a reception given by one of the society leaders in the town and here's how he wrote an account of the event for his paper.

"A woman giving the name of Mrs. J. C. Jones, who is reported to being one of the society leaders in this section, is said to have given what is purported to be a reception yesterday afternoon. It is understood that a considerable number of so-called guests were present, and some of them were quoted as saying they enjoyed the occasion. It is charged that the firm of Spivens and Spivens furnished the refreshments, and Stringem the alleged music. The hostess is said to have worn a necklace of alleged pearls which she declares were given to her by her reputed husband." [1]

[1] *The Kablegram* (February, 1939). Reprinted in *Curriculum Bulletin*, No. 100, December 20, 1951. Eugene, Ore.: University of Oregon Press.

The author of this book feels strongly that the researcher has an obligation to evaluate his material and give the reader his conclusions, generalizations, or solutions. If the researcher believes that his conclusions point to developments that are against the best interests of society, he should say so. However, he should indicate that this is a value judgment and give the reader the criteria he used in his evaluation. The reader should be left in no doubt as to where the interpretation of the evidence begins and where it ends.

The summary and conclusion section may properly indicate topics which need further research. For example, if an author has concluded that certain educational techniques are successful when the sample used is relatively intelligent and sophisticated, he may pose the following question which remains to be investigated: "Would the same technique be as successful with a sample having a much larger proportion of persons with low IQs?" When explaining the limitations of his conclusions, the writer may also suggest additional research projects. Or he may indicate that the procedures he used in finding the solution to a problem do not cover all possible situations so that further research is still needed to fill in the gaps. Sometimes more extensive sampling may lead to somewhat different answers. Or if his study is only exploratory in nature, he should indicate that extensive testing is needed before conclusions can be drawn. He should always remember that the conduct of any research investigation is similar to the opening of Pandora's box; in solving one problem he may expose a large number of other problems which need to be researched. These problems should be indicated.

If recommendations for future action grow out of the conclusions, they should be clearly stated in the report. Usually the researcher is in a better position than anyone else to make action recommendations since he is the person who conducted the study and produced whatever new evidence there is about the problem. However, he must avoid giving opinionated statements unsupported by evidence. All suggestions for further research and recommendations for action should be clearly

indicated as being the contributions of the author, and they should center about the problem being investigated.

FORMS USED IN PRESENTING EVIDENCE

The usual forms adopted for presenting the evidence obtained in research are the textual, tabular, and graphical. The *textual* form is a running narrative of the facts in an explanatory or descriptive manner. Although most of a report will be in this form, a sincere effort should be made to minimize it to avoid tiresome repetition. If the data are quantitative, every effort should be made to place them in the form of tables or graphs. This is because a textual presentation may easily become boring to a reader and thus reduce the effectiveness of the report. General information presented alone, unique material, specific references to relationships, and many other aspects of data presentation must be in text form, although long excerpts from authors whose views are relevant to the study should be placed in an appendix.

The *tabular* form is used for the presentation of quantitative data. Tables should be complete and meaningful in themselves and should not require a textual explanation. A *formal* table should be confined to a single page if possible. It should be numbered with an Arabic numeral and be given a descriptive title. The source of the data should be given. The researcher should avoid complex and unwieldy tables. If the table at first seems too large for one page, the researcher can usually reconstitute his material so that he can divide the data into two or more tables of satisfactory size. Any single table should constitute a logical unit of evidence and ought to be placed in the report close to the appropriate text discussion. Since it is undesirable to force a reader to refer forward or backward for several pages to look at a table, the writer should assemble all the text material relating to a given table in one spot, if at all possible, rather than spreading it over many pages or more than one chapter.

An *informal* table is one which is integrated into the text discussion and is not given a number and often not even a title

This form is useful for a brief classification of data which does not occupy more than six or eight lines. Various forms of formal and informal tables are discussed in Appendix A.

The *graphic* form often facilitates the presentation of data in which changes over time are stressed. It is also a good way of making clear the relative importance of subdivisions of the data, especially if time trends are important. Line graphs, maps with dots indicating areal distributions, pictographs, etc., may supplement either text or tabular presentation or may themselves show a relationship which needs no further explanation. The major advantages of graphic presentation are that it appeals to the reader's visual sense and permits a complicated set of relationships to be readily interpreted. The same general rules of construction apply to charts of various kinds as to tables. They should generally be confined to a single page, they should carry a number, and they should have both a descriptive title and a statement of the source of the data.

WRITING, REVISING, AND REWRITING THE REPORT

If the report deals with research based on quantitative data, the first draft of the report may well begin by the researcher developing tables and graphs for the presentation of his material. As each is completed, he should write out the accompanying discussion and clip the pages to the corresponding table or figure. If the research is not quantitative, he should follow a parallel procedure of gathering the cards he has prepared relating to the comments and explanations of others on the particular topic, write out his discussion, and place it with the corresponding cards.

An outline should be prepared showing the order in which the various topics will be presented. It is often desirable to write up each point on a separate page, or group of pages, even though in certain instances an entire point will require only a small part of a single page. The advantage of this procedure is that it enables the writer to reorganize the order of his presentation without cutting and pasting or retyping.

Rarely will the first draft of a written report be of acceptable

quality for an advanced academic degree at an institution of higher education. Careful revision makes the difference between a mediocre and a good piece of writing, and this process usually requires more time than the preparation of the original draft. Probably the most important and difficult part of all formal writing is the revising and reworking of the first draft. A writer finds it difficult to read his own work with the detachment and the coldly appraising eye of a reader who is going over it for the first time. A good procedure is for the writer to lay the first draft aside for a few days after it has been completed so that when he picks it up again he will be somewhat more in the position of a normal reader. If he tries to revise it too soon, he will be unaware of gaps and errors that he made originally. If he can work on some other aspect of the report for at least a week or two, he can come back and reread his own production a little more objectively. The fact that the writer is so very familiar with his topic is oftentimes a major obstacle to him in appraising his presentation from a reader's point of view. Therefore, it is often desirable to have a friend read it over critically and ask questions which will show where greater detail or clarity is needed.

As already pointed out, the writing process may expose to the author errors in his reasoning and weaknesses in his data. The revision stage may show him the need for still further research, enable him to discover erroneous conclusions, and improve the sequence of his presentation. Before starting on what is hoped will be the final copy of the report, it should be reviewed carefully for logical development, unity, cohesion, smooth transitions from idea to idea, consistency of expression, accuracy, and adequacy of the data. The criteria for judging the quality of research reports, the topic of the following section, should be used to check the report before it is considered to be in final form.

CRITERIA FOR JUDGING RESEARCH REPORTS

Few theses have been undertaken sincerely without having some strong points when they are completed even though in their entirety they are poor. Likewise, there are few good theses

that do not possess some basic weaknesses. It may be helpful to the writer of a research report—especially a thesis—to be able to measure his work against fairly generally accepted criteria so that he may be able to detect both the strong and the weak points of his report. It must be confessed that what constitutes a good thesis is not agreed upon uniformly by thesis advisors and thesis examining committees. However, the following is a digest of suggestions by a number of authors and is also something of a summary of the preceding portions of this book:

Title

1. The title ought to be brief, but still suggest the problem that the writer has chosen.

Problem Description

1. The problem ought to be one worthy of the researcher's efforts. It must be clearly explained and delimited. Some historical background as well as possible implications of the problem should be given to permit the reader to judge its importance.
2. To understand the problem better and to attempt a solution more intelligently, it should be broken down into subproblems or questions.
3. Previous works which are relevant to the problem at hand should be reviewed and appraised.
4. This section on the description of the problem is to be written in such a way that it constitutes a definite statement of the purpose of the study.

Methodology

1. The procedure by which the research was conducted should be adequately explained. Any experimental design or statistical treatment of quantitative data used must be explained.
2. The methodology must be appropriate to the problem and to the data. The reasons why the researcher chose the particular techniques he did ought to be fully explained because there often are alternatives available.
3. The data must be collected carefully and bias avoided.

Treatment of Data

1. The data used are to be accurate and adequate for the purposes of the study.

2. The data presented must be relevant; unimportant material must not appear.
3. The data are to be pertinent to a logical presentation and not be a mere collection of facts.
4. The presentation has to be objective and unbiased.
5. The analytical procedure must lead toward a solution.
6. The reasoning ought to be clear and logical. If it utilizes equations, they must be set up in a manner appropriate to the data and the problem.
7. Any standards of comparison that are utilized are to be valid.
8. Any interpretation of the data at this point must be pertinent, and if interpretations are made implications should be included also.

Summary and Conclusions

1. This section can not introduce new material. Its purpose is to digest the previous presentation and state conclusions, evaluations, generalizations, and suggestions for action.
2. The summary portions will briefly but adequately describe the problem and this particular effort at its solution.
3. All conclusions must be based on data which have been fully presented.
4. The conclusions must be based on evidence only; they are not to be merely the writer's opinion.
5. Limitations on the degree of generalization which the conclusions permit should be clearly stated.
6. Any recommendations should be made judiciously.
7. Topics for further research ought to be included.

General

1. The report should be clear and logical when viewed as an entity.
2. The format should be such that it contributes to holding the reader's attention and makes it easy to follow the discussion.
3. The style of writing should be precise, simple, and direct.

SELECTED REFERENCES

Dugdale, Kathleen. *A Manual of Form for Theses and Term Reports.* Bloomington, Ind.: Indiana University, 1950. vi + 58 pp.

Halpin, Andrew W. "Jargon—Sacred and Profane," *Phi Delta*

*Kappan, 43:*237–239, March, 1952. (Let's write to be understood!)

Hilbish, Florence M. A. *The Research Report.* New York: Bookman Associates, 1952, 292 pp.

McIntyre, Kenneth E. "The Art of Obfuscation," *Phi Delta Kappan, 43:*125–127, December, 1961. (Beclouding meaning through verbiage.)

Myers, John H. *Statistical Presentation.* Paterson, N. J.: Littlefield, Adams & Co., 1950. iv + 68 pp. (Refer to this publication for excellent illustrations and devices for presenting tables and graphs.)

Orcutt, William D. *The Desk Reference Book.* New York: Frederick A. Stokes, 1926. (Refer to pp. 160–193 for methods of indexing.)

Shannon, J. R. "Art in Writing for Educational Periodicals: The Ending," *Journal of Educational Research* (January, 1953), *46:*333–345. (Excellent treatment of the manner of writing summary and conclusion sections of a research report.)

Turabian, Kate L. *A Manual for Writers of Dissertations.* Chicago: University of Chicago Bookstore, 1937. vi + 61 pp.

University of Chicago. *A Manual of Style.* Revised and enlarged. Chicago: University of Chicago Press, 1949. x + 498 pp.

Warren, L. H. "Practical Suggestions for Reducing the Labor of Indexing a Textbook," *Science* (September, 1940), *92:*217–218.

Wheeler, Martha Thorne. *Indexing; Principles, Rules and Examples.* 3d ed., Albany, N. Y.: New York State Library School, 1923.

APPENDIXES

APPENDIX A

A Form Manual for Academic Research Reports

Format
- Title page
- Approval sheet
- Acknowledgment and preface
- Table of contents and lists of tables, figures, and plates
- Context chapters
- Bibliography, appendix, and index

Construction and Placement of Tables and Figures
- Layout on the page
- Title and number
- Unit of measure
- Captions and stubs
- Footnotes
- Spacing and ruling
- Use of zero
- Classification and order of arrangement
- Mechanical aids

Reference to Sources
- Bibliography and footnote style
- Indices for footnotes and bibliographical entries
- Abbreviations used in footnotes

Typing and Miscellaneous Considerations
- Number of copies of theses and abstracts

Time for submitting reports
Materials
Margins and spacing
Page numbering
Inclusion of illustrative material
Spelling and division of words
Use of numbers
Abbreviations
Punctuation
Responsibility

Example Pages

If a researcher will examine a number of professional journals and books, he will recognize that there is no universal agreement about the proper form for headings, footnotes, tables, figures, graphs, and the like. If he is preparing his manuscript for publication, he must find out in detail just what is required by the editors of the publications in which his material is to appear. If he is preparing to write a term paper, field study, thesis, or dissertation, he will find that although each institution maintains some standard of form for theses and dissertations these standards differ from institution to institution.

Many institutions of higher education recommend that writers of research reports follow one of the published form manuals, of which several are available.[1] One of the difficulties students encounter in the use of some of these manuals is that many times the forms recommended are intended for printed publications rather than for material which is to be produced in typewritten form. The recommendations in this appendix are made for material that is to be submitted in typewritten form in partial fulfillment of degree requirements. While there are several "correct" ways to organize and present data, a mixture of several forms would lead to confused copy. Therefore,

[1] Probably the best known is *A Manual of Style,* 11th ed. Chicago: University of Chicago Press, 1949, 534 pp.

this appendix presents *one* form that is usually acceptable for term papers, field studies, theses, and dissertations.

FORMAT

The logical divisions of a good outline for a report of a research study are customarily arranged in the following sequence:

- Title page
- Approval sheet
- Preface (if any)
- Table of contents
- List of tables (if any)
- List of figures (if any)
- Context chapters
- Bibliography
- Appendix (if any)
- Index (if any)

Title Page

The title page should carry a concise and adequately descriptive title of the research study. Its principal use is to tell a prospective reader whether or not to refer to the report. Thus, it should express as much of the scope of the report as possible and indicate the general method employed. It should include expressive key words since many indexes to research reports are made in terms of such key words. However, it should avoid the use of vague generalities, superfluous words, or the type of seductive captions currently used for novels and the movies.

The entire title should be typed in *all capital letters* and, if more than one line is required, it should be broken and arranged to make a pleasing appearance on the page. No closing punctuation should be used in the title. The title page should also include a statement with respect to the academic degree for which the degree is conferred. An example of a title page is shown on p. 300.

Approval Sheet

The approval page should follow the title page and provide a line upon which the faculty advisor (and members of a faculty committee, if customary in a particular institution) may sign

his name. The advisor's name should be typed below the line. An example of an approval sheet is shown on p. 301.

Acknowledgment and Preface

A page acknowledging the help of others in the research study is often included immediately following the approval sheet (see p. 302). The comments should be brief, simple, temperate, tactful, and modest, and given only for substantial assistance and cooperation of a nonroutine character which warrants public recognition. Strong and valid criticisms of many acknowledgments by graduate students have often been made, of which the following excerpt from a publication of the University of Cincinnati is an example:

> Bestowal of praise or expression of indebtedness in effusive, sentimental, or extravagant language may cause the reader to question the sincerity of the investigator or the soundness of the study. The listing of well-known names, whether on the faculty of the local institution or elsewhere, to court favor or to enhance the value of the study by lending to it a fictitious authority, is a form of intellectual and professional dishonesty. Types of acknowledgment that have become stereotyped and may well be avoided are those referring to: the patience and tolerance of a spouse during the pursuit of the graduate degree, exaggerated tributes to the advisory committee, contributions of the graduate classes attended, librarians from whom books were borrowed or references secured, efforts of the typist, clerks assisting in the scoring of tests or tabulation of data, operators of calculating machines, and casual or occasional interviews or letters.[2]

A preface is not usually included in a graduate student's research report unless it is necessary to explain more about the history, scope, methodology, or the writer's opinion of the place of the study in the field of knowledge than can be properly given in the introductory chapter. The preface, when used, is usually signed or initialed by its writer.

[2] *A Guide for the Preparation of Dissertations, Theses, and Field Reports.* Cincinnati, Ohio: The Faculty of Teachers College, University of Cincinnati, 1952, p. 14.

Table of Contents and Lists of Tables, Figures, and Plates

The purpose of the table of contents is to provide a synopsis of the design or pattern of the research report. If logical and appropriate topical headings are used in the report, the table of contents presenting them becomes helpful to a reader. The table of contents may contain only a list of the titles of chapters and their appropriate Roman numerals, followed by the page number on which each chapter begins, or it may be more analytical, containing, besides the chapter titles, subheads or section headings or words or phrases indicating the subject matter of the chapters. Subtitles, however, should be indented under each chapter title and be followed by the specific page or section references. If reference to specific pages for subtitles is not desired, the subtitles may be run together and separated by semicolons or dashes. An example of an acceptable table of contents is shown on pp. 303–304.

Separate pages for the list of tables and figures should follow the table of contents. Both tables and figures sometimes are allocated an entire page; where this is done for figures the words "plate" or "illustration" are sometimes used. These lists should contain the number of each, its exact title, and the page where it is found in the report. These should be listed in the order in which they appear in the report, and consecutive numbers should be used for each. Arabic numerals are usually used for identifying tables, figures, and plates. Tables are usually limited to the presentation of verbal or numerical data, while figures and plates may include all types of illustrations such as graphs, charts, diagrams, maps, and photographs. In the lists of tables and figures, the titles should be typed with initial capitals, rather than all capital letters, on all important words. However, some institutions and some publishers prescribe all capital letters for all titles and captions in tables and figures. Whatever form is used should be used consistently throughout a research report. Examples of these lists are shown on pp. 305 and 306.

Context Chapters

The context chapters should be organized into at least three logical parts: (1) an introduction, (2) a presentation of evidence for the solution of the problem presented, and (3) a summary and conclusion.

The Introduction. The introduction should include a statement of the nature and importance of the general problem of the study. This should be made manifest to a reader within the first few pages of the report. It may also contain a review of the pertinent literature and a discussion of the various aspects of the problem. The major parts of the problem should be delineated, and there should be a brief discussion of the materials and procedures used in their solution. It may be desirable to include a brief overview of the nature of the contents of the main body of the thesis.

Presentation of Evidence. Each major division of the problem should be presented in a separate chapter. The chapter should include a discussion of the issue or part of the problem investigated and the evidence used in its solution. If this becomes lengthy, a summary of the evidence may be made at the end of the chapter.

Each chapter should begin on a new page, and a chapter Roman numeral and title should be given. The title should provide a short, concise description of the chapter content. It is not usually desirable to use general operational descriptions of chapter titles, such as "Procedures," "Methods of Research," "Collection of Data," and "The Findings." It is better to use more descriptive titles, such as "Professional Competencies Basic to Oral Communication Training of Teachers," or "Analysis of Performance Ratings."

If a chapter is very long it may be desirable to subdivide it into two or more chapters, repeating the original heading for each of such chapters and adding, within parentheses, the word "continued."

The subdivisions of the major elements of a chapter should be emphasized by the use of *headings*. These headings may be

thought of as being parallel to the outline for the study and should be used as follows:

CHAPTERS: Major problems of the research project.
Center headings: Major topics within each chapter.
Marginal headings: Minor topics related to each of the major topics.
Paragraph headings: Subdivisions of the minor topics.

A single subdivision heading is not desirable. Unless there are two or more breakdowns within a level, the previously used higher level heading should be sufficient. It is not necessary to use headings for all material, since introductory or transitional paragraphs may precede a heading and summary paragraphs may follow headed material. However, a writer should be consistent in whatever plan he uses. An example page showing headings is given on p. 308.

The chapter titles should be typed in *all capital letters.* If subordinate headings are used, the following procedures of typing and underscoring should be observed.

Center and marginal headings. Use initial capitalization on all important words, and underscore all words. No period is used at the end of either of these types of headings.

Indented paragraph headings. The first word of paragraph headings should be capitalized, all words underscored, and the headings should be followed by a period. The context material continues on the same line as the heading, as shown in this particular paragraph.

A *footnote* presenting only a bibliographic reference should list the reference exactly as it is given in the bibliography except that the specific page or pages referred to are listed instead of the total pagination of the reference. Explanatory footnotes are often used to avoid breaking the continuity of thought in the body of the text and to provide cross references to other parts of a chapter or of the report. Further details with respect to footnotes are presented in the section on *reference to sources* on pp. 288–293.

Summary and Conclusions. One or more chapters may be used to present a summary of the study, the conclusions derived, their implications, and the writer's recommendations. This final section should include an overview of the entire study which may be of particular value to those readers who do not wish to go through the entire dissertation for information concerning the problems, treated sources, methods, and conclusions.

The section should begin with a statement of purpose and the formulation of working hypotheses. Since the "presentation of evidence" chapters have included the evidence and conclusions derived from it with respect to each element of the problem, the final chapter should only recapitulate the answers to the initial questions or problems. The conclusions do not need to repeat the evidence on which they are based, but extreme care should be exercised to present them with whatever limitations or qualifications are necessary.

It is desirable to include a section dealing with implications of the study and recommendations of the researcher with respect to their applicability. Some authorities on research reporting do not agree with the author that a researcher should "go beyond" the presentation of evidence and the drawing of conclusions based on that evidence. However, it should be remembered that the researcher is the one who has conceived the study, has broken it down into its constituent parts, has collected the evidence in terms of his understanding of the problem, and has reached conclusions on the basis of the evidence he has found, interpreted, and presented. Therefore, he is in a better position than anyone else to know from what point of view the problem has been attacked and should have not only the freedom, but also the responsibility, to present his interpretations of the evidence and his recommendations with respect to its use. It should also be remembered that any reader of the research report may have divergent opinions with respect to these implications and recommendations in terms of his own understanding of the problems, evidence, and conclusions presented. The inclusion of the researcher's recommendations may

serve as a check on a reader's interpretation of the report, and may prevent a reader from making erroneous interpretations as the result of cursory reading.

Bibliography, Appendix, and Index

Bibliography. Any report representing extensive research should contain a bibliography for completeness and convenience of reference. In graduate student theses and dissertations all sources of references presented in footnotes should be included in the bibliography. In publications, however, publishers sometimes object to this duplication as an expense, if the footnotes are complete, and recommend that the bibliographical entries should include only those in addition to footnotes. The main reason for including footnote references in the bibliographies of theses and dissertations is to provide a source of all references for a topic in one place.

In addition, it is sometimes desirable to include certain materials not previously referred to that relate directly to the particular problem. Bibliographic references may be briefly annotated to make the bibliography more meaningful, especially if no review of them has been presented earlier in the report. The references should be numbered consecutively if a cross-reference system of citation from the body of the report to the bibliography is used in place of footnotes. Otherwise, numbering is unnecessary, but may be used. Bibliographical materials may be classified into sections as books, periodicals, and so forth, and if the bibliography is extensive, it may be desirable to list each classification under separate headings, such as (1) Books, (2) Periodicals, (3) Documents, (4) Reports, and (5) Unpublished materials. The classifications adopted depend largely upon the character of the materials and the nature of the research problem. Within each classification, the bibliographic materials should be arranged alphabetically by authors, or by titles if the work is anonymous or edited. Examples of bibliographic form are presented on pp. 313–319.

Appendix. An appendix is used for additional or supplementary materials, such as statistical tables, schedules, ques-

tionnaires and interview forms used in the study, documents, and long explanatory notes to the text. All material not essential to an understanding of the text, but useful as supporting evidence, should be placed in an appendix. However, an appendix should not be included if it can be avoided. If included, each group of data or material should be given a title and referred to by title in the body of the report. If more than one appendix is necessary, or if the appendix is divided into sections, each part should be designated by a capital letter, rather than a number.

Index. An index is not usually included in graduate student research reports. However, if the report is being prepared for publication and is a technical manuscript or is intended as a work of reference, an index is desirable. If an analytical table of contents is used, the index is less necessary, but desirable, as it makes the work more usable. The index should give an alphabetically arranged, detailed reference to all important matters discussed in the report, such as names of persons, places, events, titles of books, definitions, and vital statements. Since an index is seldom used in graduate student research reports, a detailed discussion is unnecessary at this time. The procedures and techniques of indexing are available in several sources.[3]

CONSTRUCTION AND PLACEMENT OF TABLES AND FIGURES

A basic principle to follow in the construction of tables and figures is to make them as attractive and simple as possible and still present the data in such a way as to make the evidence manifest to a reader. Tables and figures should be placed in a manuscript as nearly following the point of first reference as possible. Only those tables or figures that present information essential to an understanding of the manuscript should be included in the body chapters. Supplementary tables and figures

[3] Walter V. Bingham, "How to Make a Useful Index," *American Psychologist*, January, 1951; pp. 31–34. William D. Orcutt, *The Desk Reference Book*, New York: Frederick A. Stokes, 1926, pp. 160–193; L. H. Warren, "Practical Suggestions for Reducing the Labor of Indexing a Textbook," *Science*, September 6, 1940, pp. 217–218; and Martha T. Wheeler, *Indexing: Principles, Rules and Examples*, Albany, N.Y.: New York State Library School, 1923.

which may aid a reader in making a more detailed study of the report should be placed in an appendix. It is usually desirable to include long, highly detailed tables in an appendix, using shorter, summary tables in the body of the report. Tables and figures should be so constructed as to be understandable apart from the text of the report. The principal points to be considered in the construction of tables and figures, in addition to the form of ruling, spacing, structure, and capitalization, are shown in the examples on pp. 309–312.

Layout on the Page

A table may appear in any one of three ways. Types of tabular material that are brief may be inserted directly in the context of a report. In this case the tabulated material is usually introduced by a sentence followed by a colon and presented single-spaced in the same manner as lengthy quotations or lists of items. The following paragraph illustrates the use of an *informal table* in context.

Data from the total norm group were used in the computation of norms. The tables of percentile norms were based on a number of pupils in each grade as follows:

4th grade–1st semester–1344
4th grade–2nd semester–1421
5th grade–1st semester–1360
5th grade–2nd semester–1389
6th grade–1st semester–1463
6th grade–2nd semester–1432

A more *formal setup* might be used for a somewhat larger amount of data, as illustrated below.

Composition of the Total Norm Group

Grade	Semester	Median Age	Number of Pupils
4	1	9 yr. 5 mo.	1344
4	2	9 yr. 11 mo.	1421
5	1	10 yr. 5 mo.	1360
5	2	10 yr. 11 mo.	1389
6	1	11 yr. 5 mo.	1463
6	2	11 yr. 11 mo.	1432

A third method is to place the table on a separate page. That procedure could have been used with the previous example, but since the amount of material was so small, it would have been a waste of space and might have tended to place more emphasis upon the material than is warranted by the number of items. However, in writing a research report it is desirable to place all tables on separate pages to avoid recopying them during the process of writing and revising the research report. After the report is in final form and ready to be copied for presentation, informal tables and formal tables taking up no more than approximately one-third of a page each may be included in context. The longer tables should always be presented on separate pages. It is usually desirable to present *all* figures on separate pages rather than to incorporate them with pages containing textual material.

If a table is presented to accompany a figure, it should be adjacent, usually preceding, and subordinate to the figure. In many instances both may occupy the same page. If a figure occupies a full page, the corresponding table should be placed on the opposite page facing the figure, and the front of the page should be left blank except for the page number. In this manner a reader may readily refer to the table without the inconvenience of turning the page.

It is preferable to place tables or figures on a page in such a manner that the pages does not need to be rotated to look at them. Where it is necessary, on account of the length of rows in a table or the shape of a figure, the table or figure may be set lengthwise on the page. In this case the rule is to place the bottom of the table or figure at the *right* side of the page. If the table or figure is placed sidewise, place the lettering or captions so that they may be read from the side or bottom, but do not place them so that the book needs to be rotated more than once.

Title and Number

Every table or figure must have a title that is complete, concise, and descriptive. The title should answer, in general, the

questions relating to who, what, where, and how many. For a table, the title should be written in initial capitals and is usually placed *at the top* of the table. The title for a figure is usually placed *below* the figure. The word "table" or "figure" and its number should be placed two spaces above the title and written in initial capitals.

Unit of Measure

The unit of measure for items in tables and figures should always be indicated prominently, usually in parentheses, below the title. It should never be subordinated to a footnote, except in cases where different units are used for different items of a composite table. For example, a title may be expressed as follows with the unit of measure included within parentheses:

Federal Government
Budget Expenditures and Revenues
Fiscal Years, 1947–1961
(Millions of Dollars)

Captions and Stubs

The captions (column headings) and stubs (labels for rows) must tell exactly what is contained in the data presented and yet should not be so long as to detract from the conciseness of a table or figure. Common abbreviations may be used, but unfamiliar ones should be explained in footnotes at the bottom of the data presented. If possible, all abbreviations should be avoided. It is not desirable to use code numbers in connection with a table or figure unless a list of the items and code numbers is presented on an adjacent facing page. The best practice is to use a brief description in a caption or stub with a footnote for further explanation.

Footnotes

It is desirable to use letters or typewriter characters for indicating footnotes in tables and figures to avoid the possibility

of confusing the index with the numerical data presented. Footnotes should be placed below the horizontal ruling at the bottom of a table. The source of data for a table or figure should be given in a footnote if the source is other than the researcher's.

Spacing and Ruling

The major purposes of spacing and ruling are to aid a reader in following a long line of data and to help him to break the data into logical groups. In complex tables using several columns and rows it is sometimes desirable to double-space at strategic points, usually every fifth row. Horizontal and vertical rulings should be kept to a minimum. Usually a double horizontal ruling is placed between the table title and the captions, a single horizontal ruling is placed below the captions and above the first row in the series or items, and another horizontal ruling is usually placed below the last row in the series of items. Vertical rulings, and other horizontal rulings, should be included only as needed in making the table more easily understood. Rulings are not necessary on the margins of tables, but they are usually included on figures.

In graphical presentations the figures are usually drawn on plain paper with only a few grid lines to guide a reader. A researcher may use highly detailed graph paper in the original construction of his figure, but should transfer the figure to plain paper with only a few grid lines for the final presentation. Since the grid lines should be background material, they should be drawn lightly and pass back of any lines, bars, or figures presented.

The lines used in various figures may differ in size and form. The essential parts of a figure should be indicated by the heaviest lines and subordinate aspects should be indicated by lighter lines. It is never desirable to use colored lines in research reports submitted in the fulfillment of graduate degrees. Colored lines in research reports are expensive to reproduce, and if any of the work is to be microfilmed or reproduced by a direct black-and-white photographic process for later publication, the

different colors may not appear in the same tone of black. In addition, if the reproduction is accurate, the figure may then be confusing since all lines would be, approximately, the same tone of black. It is better to use a variety of solid, dotted, broken, and similar types of lines to avoid any confusion after reproduction. Furthermore, some readers may perceive different colors with different emphasis that is not warranted by the data nor intended by the researcher.

Use of Zero

The zero value should not be used in tabular presentations to indicate a *lack* of information. It may be used only when it represents a *value* of zero. When there is a lack of available information, the letters NA (not available) may be used or the space may be filled with dots (. . .) or some other characteristic explained in a footnote indicating that no information was available.

Classification and Order of Arrangement

Classification refers to the subdivisions presented in a table or to the scales used in figures. It is desirable to select for a table or figure the data needed to establish clearly a limited number of points with only a few groups of columns or rows of data. However, if longer tables are necessary, it is preferable to break them into two parts with a second part continued to another page, rather than to place them on pages that have to be folded. When a table is continued to a second page, all the headings should be repeated, and the table or figure number, but not the title, should be repeated and followed by the word "continued." The following rules should be used in classifying and arranging the data:

1. Have enough classes to condense the data sufficiently but not to the point of losing significant features of a distribution. The general rules for grouping data for frequency distributions apply in this case.
2. Data should be presented in a logical order of progression. For

example, frequencies should progress from high to low, or vice versa.

3. Uniform class intervals are desirable to aid in interpretation. The class limits must be clearly and precisely stated. Midpoints of class intervals may be used if there is no doubt as to the size of class limits. It is most convenient if interval midpoints can be whole numbers.
4. When line graphs, histograms, frequency polygons, or ogives are presented with an arithmetic *vertical* scale, the zero should appear. If the zero would not ordinarily appear without having an excessively long figure, it should be shown by the use of a horizontal break in the diagram. If the vertical scale is logarithmic, zero cannot be shown, and if it refers to percentages, it need not be given.
5. The zero point does not need to be included in the *horizontal* scale of a figure.
6. Figures for the scales of a diagram should be placed at the left and at the bottom or along the respective axes.
7. It is often desirable to include in a figure a table showing the numerical data given in the graph.

Mechanical Aids

Rulings for tables and graphs may be made on a typewriter, but if they are hand-drawn they should be drawn with India ink. Titles and captions may be typed or hand-lettered. The use of a mechanical lettering device, of which many different types are available, is helpful in producing an attractive and easily read figure. For cross-hatching and other shading work the use of specially printed, gummed cellophane designs speeds the preparation of a neat graph. Many usable devices are available at most draftsman's supply shops.

REFERENCE TO SOURCES

As a general rule research reports should give credit for the source of information by a footnote rather than as a part of the text discussion. Documentation not only provides help for the reader who wishes to explore a point in greater detail, but also may strengthen the report because of the prestige of the work

cited. Many times cross references to other parts of the report aid in avoiding unnecessary repetition or breaks in the continuity of thought in the body of the text.

At the end of the research report all sources referred to are listed alphabetically by the surname of authors. It is desirable in terms of completeness and convenience of reference to include in the bibliography all the items to which footnotes refer.

Bibliography and Footnote Style

Bibliographical entries and footnotes referring to other sources of information should be made in a form to facilitate identification and location of the materials cited. In general, each entry should consist of the author's name, title of the material cited, place of publication, name of publisher, date of publication, and pagination. Several styles are used by various institutions and publishers, but there is general agreement that within any single publication the style should be consistent.

Author's Name. In a bibliography the author's name should be given with surname listed first, separated from the full given name (if known), and followed by a period. Semicolons may be used between names when there are three or more authors. In a footnote, the name of the author may appear in normal order.

Title of Material Cited. The title of a part of a publication, such as a chapter in a book or an article in a journal, should be placed within quotation marks and separated from the title of the entire publication by a comma. The title of the entire publication should be underlined and followed by a period. If there are any other facts essential to identification of an entry, such as editor, edition, number in a series, introduction, or translation, they should come immediately after the title with each item followed by a period.

Place of Publication. The date used should be the one found on the title page of the publication and should be followed by a period. If no date is given on the title page, the copyright date should be given.

Pagination. In footnote usage the specific pages referred to are entered following *pp.* (plural) or *p.* (singular) and fol-

lowed by a period. In a bibliography the total pagination including the front matter and the text may be given. If this is done, the front matter is given in small Roman numerals followed by a plus sign, the total number of pages in Arabic numerals, followed by *pp* and a period. If certain specifically designated pages are given in the bibliography, the title should be followed by a comma and the number of pages. In this case the total pagination is given also as above.

The following are examples of acceptable form for footnotes and bibliographical references. The first is for a book, the second for a chapter in a book, and the third for a periodical. Additional samples illustrating various types of references are shown on the example pages at the end of this chapter.

[1] Good, Carter V., and Scates, Douglas E. *Methods of Research.* New York: Appleton-Century-Crofts, 1954. xx + 920 pp. (Or pp. 865–866, if used in footnotes for specific reference.)

[2] Good, Carter V., and Scates, Douglas E. "The Reporting and Implementing of Research," *Methods of Research,* pp. 832–896. New York: Appleton-Century-Crofts, 1954. xx + 920 pp.

[3] Burton, Willam Henry. "Remarks Upon Academic Freedom," *Educational Leadership,* 10:505–507, May, 1953.

Note that in a reference to a periodical, the first numeral given is the number of the volume and the following numerals show the pages cited. In footnote usage the pages listed are those to which particular reference is made rather than the total pagination of the book.

Some style manuals make a distinction between the form of reference to an author's name in a footnote and in a bibliographical entry. Some state that for a footnote the author's name should be written in normal order (e.g., Carter V. Good and Douglas E. Scates in the example previously given), while others use the same form for both footnotes and bibliographical entries. The use of the same form tends to simplify the procedure in typing footnotes and bibliographical entries from bibliography cards. However, either form is acceptable if used consistently.

Indices for Footnotes and Bibliographical Entries

Reference to a source that is given in a footnote is by a numerical index. This index is usually placed after the key noun or major statement in the material or idea that is being documented. When the index comes at the end of a phrase or sentence, it should be placed outside the punctuation mark.

The numerical index should be raised a half-space in text and in the footnote. In typewritten research reports footnotes should be numbered consecutively on each page, beginning with the Arabic numeral 1. However, some style manuals recommend that the numbering continue throughout each chapter of a report, beginning the renumbering anew with each chapter, while others continue the numbering throughout the entire report. The author recommends consecutive numbering for each page, only, to avoid the necessity of renumbering if it proves necessary to insert one or more citations after the original writing. In this manner each page of the report is independent of other pages. If a researcher plans at a later date to publish his report in a monograph or journal which necessitates a change in the style, such changes as might be necessary are more readily made from this form.

In such publications as the *Review of Educational Research, The Journal of Experimental Education,* the *Encyclopedia of Educational Research,* and some others, the index number is written in parentheses in alignment with other materials on the page. For example, an index such as (3) following an author's name would refer to item number 3 in the bibliography. If the index were (3:19–22) or (3,19–22) following an author's name, it would mean that pages 19 to 22 of item number 3 in the bibliography contain the material cited. The use of this method necessitates a complete bibliography consecutively numbered at the time a researcher begins his report. While this may be preferable as a style for publications, since it saves space by reducing the amount taken up by footnotes, the former style of using the numerical superscript is preferable for graduate student research reports.

Footnotes included in figures and tables, especially those containing other materials in numerical form, may be somewhat confusing if the numerical superscripts are used. To avoid confusion, it is acceptable practice to use an asterisk (*), or other common typewriter characteristics (#, §, !) instead of numerals in tables, formulas, or mathematical materials.

The footnotes should be separated from the body of a page by a short line, usually of 20 spaces, from the left-hand margin.[4] The footnotes should be single-spaced within a footnote, with a double space between footnotes. They may be set in block form beginning at the left-hand margin, or paragraphed. The footnote index should precede the footnote and be raised a half-space.

Abbreviations Used in Footnotes

When the same sources are referred to several times, it is permissible to use certain abbreviations to avoid repeating the full reference each time in the footnotes. However, abbreviations should not be used to refer to citations in previous chapters. The following abbreviations are commonly used.

Ibid. This is an abbreviation of *ibidem,* which means "in the same place." It may be used when succeeding uninterrupted citations of a work occur *on the same page* of a report, and should not refer to a citation made on a different page. It is appropriate to cite the work completely when first used on a page.

Op. Cit. This is an abbreviation for *opere citato,* which means "in the work cited." This is the proper abbreviation to use when other references intervene between different citations of a particular work. When it is used, the author's name is given first, followed by *op. cit.,* and the appropriate page numbers.

Although there are many more abbreviations that can be

[4] While this method is widely used, it is sometimes more convenient when preparing a manuscript for printed publication to "run in" a footnote in the line immediately following the index for it, but separated from this line and the one at the end of the footnote by a broken line. A writer should, however, follow the specific regulations set forth by the institution or publisher to whom his research report is to be submitted.

found in various books, it is not desirable to use any but the above for the identification of sources for graduate student research reports. Occasionally it may be necessary to refer to many different pages within a work cited. In this case the following abbreviations may be used.

Et. Seq. This is the abbreviation for *et sequences,* which means "and following." It may be used to refer to material contained on several successive pages and is usually indicated following the first page cited. The English abbreviation, ff., may be used in the same manner. However, there should be consistency in the use of either the Latin or English abbreviations.

Passim. If a writer wishes to indicate that the material referred to is scattered throughout a work, this abbreviation, meaning "here and there," may be used.

TYPING AND MISCELLANEOUS CONSIDERATIONS

There are a number of minor considerations dealing with the typing of research reports for submission to graduate schools, two of which are the number of copies and time for submission. These are discussed in the following sections.

Number of Copies of Theses and Abstracts

The number of copies of a research study which need to be submitted varies somewhat from institution to institution. A student should find out the number required by the particular institution from which he is planning to receive his degree.

Masters' Theses and Doctors' Dissertations. In general, a student should have at least five copies of a master's thesis or a doctor's dissertation for the following purposes:

2 to be deposited in the institution's library.
1 for the major department library.
1 for the faculty advisor of the study.
1 (or more) for the student's own use.

The copies for the institutional library usually consist of the

original and the first carbon copy. These are submitted unbound, but in a large manila envelope, to the graduate school.

An abstract is a digest usually not exceeding 1500 words which gives the highlights of the problem description and the conclusions. The number of copies of *abstracts* usually exceeds the number of copies of the research reports since they are sent to each of the members of the faculty committee which administers the final examination for an advanced degree. Therefore, it is usually necessary to have from seven to ten copies of the abstract to accompany the thesis or dissertation when it is turned in to the graduate school.

Field Studies and Other Research Project Reports. There is no general pattern for the submission of reports other than Masters' theses and Doctors' dissertations. Usually lesser reports are submitted only to a student's faculty advisor. Some institutions require that these be bound prior to submitting them, while others will accept them unbound. A student should type at least one copy for his advisor and have one copy for himself.

Manuscripts for Publication. When a research report is being submitted for publication in a book, professional journal, or monograph, usually only one copy is submitted. However, it is a good plan to type one or two copies in addition to the original to provide security against loss and for reference in subsequent correspondence with a publisher.

Time for Submitting Reports

It is customary practice for institutions granting graduate degrees to require the submission of reports to the graduate school prior to taking a final examination for a degree. Usually these reports need to be submitted from two to four weeks prior to the examination. After they have been checked by the graduate school, they are usually distributed to members of the final examination committee to read prior to the time scheduled for the examination.

Materials

The research report should be clearly typewritten in standard

pica (large) or elite (small) type on a good quality of paper, preferably having a rag content of not less than 50 percent. The majority of graduate schools require that the original and first carbon copy be typed on 16- or 20-pound paper. Other copies may be typed by carbon on 16-pound paper, or some thinner paper, such as onionskin. The ribbon used on the typewriter should be in good condition and black in color. A good grade of black, greaseless, and hard carbon paper should be used for copies.

Margins and Spacing

The general margins should be 1 inch wide at the top, bottom, and right-hand side of the page. The left-hand, or bound side of the page, should have a margin of 1½ inches.

On title pages, such as those carrying the chapter numbers and titles, the chapter number should be written in all capital letters (e.g., CHAPTER VI) and placed 2 inches from the top of the page. The chapter title should be centered, typed in all capital letters, and should fall two spaces below the chapter number. The first line of text should begin three spaces below the chapter title.

All straight text should be double-spaced. Quoted materials, in excess of a few lines, should be indented five spaces and typed in block form single-spaced. Lists of items, principles, tabulations, enumerations, and footnotes should be single-spaced. When quotations are single-spaced and indented, the quotation marks are omitted.

The beginnings of paragraphs should be indented five spaces. Subordinate indentations may vary in extent, but they should be consistent throughout the report.

When center headings are used, three spaces should be allowed above and below each heading. Two spaces above and below marginal headings should be allowed, and three spaces should be used above and below tabulations in context.

Page Numbering

All pages in a report should be given consecutive numbers.

The front matter, such as title page, table of contents, etc., should be numbered with lower-case Roman numerals. However, the number is omitted from the title page.

Beginning with the title page of the first chapter, all subsequent pages in the report, including the bibliography, appendix, etc., should be numbered consecutively with Arabic numerals. The number is allowed for all subsequent pages containing chapter titles, but it is not typed in.

Page numbers should be typed in the upper right-hand corner of the pages and spaced one inch from the top and side. Two spaces should be allowed below the page number for the first line of typing.

Inclusion of Illustrative Material

Photographs, maps, or other loose illustrative materials, should be pasted on the paper used for typing and placed under a press until dry to prevent wrinkling of the paper. It is necessary to use a permanent type of adhesive, rather than rubber cement, for pasting loose materials. The general rules pertaining to tables and figures apply to this type of material.

Spelling and Division of Words

Do not use simplified spelling, such as "tho" and "thru," in research reports. If there is any question as to the proper form for spelling a word, it is recommended that a recent, unabridged dictionary be used as a guide.

Avoid unnecessary division of a word at the end of a line. Words of less than four syllables should not be divided, and the conventional rules should be followed according to the natural divisions in correct pronunciation.

Use of Numbers

As a general rule, numbers below one hundred and round numbers, such as three billion or five thousand, should be spelled out. When numbers are used at the beginning of a sentence, they should be spelled out. There are a few situations

in which there may be a good reason for not following these general rules, however.

Figures should be used to express dates, street numbers, page numbers, decimals, and percentages. It is not considered good form to use the symbol % for percent except in the body of tables. When percent is used, it requires neither italics, underscoring, nor a period, but must always be preceded by a figure. It should never be used in place of the noun, percentage.

When numbers are presented for comparison, or when several numbers occur in the same paragraph, figures should be used. When two series of numbers are involved, spell out one series and express the other series in figures. For example: "Thirty schools had an average daily attendance of 200 pupils; twenty-two averaged 150 pupils; and fifteen averaged less than 100 pupils."

When two numbers occur together without punctuation, one of the numbers may be spelled out for clarity of expression. For example: "The class included 12 five-year-old pupils, 18 six-year-olds; and 3 seven-year-olds."

When numbers used in tables and figures are discussed in the text, express them in figures regardless of the general rule. Whatever manner of expressing numbers is used, there should be consistency throughout the report.

It is unnecessary to place "00" after a decimal point or before numbers beginning with a decimal point. Where a small figure or sum of money is only mentioned occasionally, it may be preferable to spell it out, otherwise use figures. For example: In an isolated case it is more appropriate to write thirty-nine cents than $0.39 or 39¢.

Abbreviations

It is undesirable to use abbreviations in research reports as they might be misunderstood. However, there are a few exceptions. All titles preceding personal names are spelled out with the exception of Mr., Mrs., and Dr., which are always abbreviated.

Punctuation

The general rules for punctuation should be applied in research reports as in other academic and scientific writing. The following rules are presented, however, for they pertain to frequent misuses among students.

Periods and commas are always placed inside quotation marks, but other marks of punctuation are placed inside or out as required by the construction or meaning of a sentence.

The hyphen is a single mark without spacing on either side; e.g., one-half.

The dash is made by two hyphens without spacing on either side; e.g., John Doe—the man of the hour—was born in . . . It is used to indicate an abrupt change of thought.

If words are left out of quoted materials, ellipsis marks are used to indicate the omission. If the omission is at the beginning, or in the middle of a sentence, simple ellipsis marks (. . .) are used. If the omission is at the end of a sentence, the simple ellipsis marks as well as a period (. . . .) are used.

If something of the writer's is added to an original quotation, such as a name, date, pronoun, or comment without which the quotation would be meaningless, the added material is enclosed in brackets, [ ].

Symbols not easily constructed with a typewriter should be inserted neatly by hand in India ink.

Responsibility

In the final analysis the writer of a research report, not the typist, is responsible for the contents of the report. He is responsible for all errors, regardless of kind. Therefore, he should accurately proofread the manuscript before submitting it and should eliminate all errors. The typist should be held responsible for correcting all typographical errors without charge. However, the research writer should pay for all errors due to his negligence or poor copy.

EXAMPLE PAGES

On the following pages appear typewritten specimen pages representing the various parts of a research report. In some cases comments concerning a page are included within brackets on the page. In referring to these pages, the research writer should bear in mind that adaptations should be made to conform to any specifications provided by the school or department of the institution in which he is working and to which he is intending to submit his report, and that there is no single "correct" form. The examples included here illustrate general practices rather than "rules" or "laws."

[Example of a Title Page]

TRAINING OF SECONDARY SCHOOL MUSIC TEACHERS IN

THE UNDERGRADUATE PROGRAMS OF COLLEGES

AND UNIVERSITIES OF SEVENTEEN

WESTERN STATES

by

WILBUR JOHN PETERSON

A THESIS [OR FIELD STUDY OR DISSERTATION]

Presented to the School of Education and the
Graduate School of the University of Oregon in
partial fulfillment
of the requirements for the degree of
Doctor of Education [Or Master of . . .]

June 1954

[Example of an Approval Sheet]

APPROVED:

____

(Advisor for the Thesis)

[Additional lines may be added for the signatures of the members of the final examination or thesis committee. A student should check this form for the particular institution in which he is enrolled.]

[Example of Acknowledgments]

ACKNOWLEDGMENTS

The writer wishes to express his appreciation for the guidance given by Dr. Robert Nye and Dr. J. Francis Rummel in the preparation of this thesis.

[Example of Table of Contents]

TABLE OF CONTENTS

[Example of Second Page of Table of Contents]

TABLE OF CONTENTS (Continued)

[Example of List of Tables]

LIST OF TABLES

[This table was continued in a similar manner.]

[Example of List of Figures]

LIST OF FIGURES

[This was a complete list of figures for a dissertation.]

[Example of a First Page of a Thesis]

CHAPTER I

THE NATURE AND PURPOSE OF THE INVESTIGATION

In educational work it is frequently desirable to excuse some individuals, on the basis of examinations, from specific requirements of an educational institution. For example, at the State University of Iowa all students are required to meet before graduation certain standards of proficiency in the areas of communication skills, mathematics skills, and physical education skills. The skills which are considered basic in mathematics have been defined by the faculty.[1] Students entering the College of Liberal Arts for the first time are given a test[2] which covers these basic skills in mathematics. Since it has been determined by the faculty that approximately 50 per cent of a typical group of entering freshmen possess these skills to a degree which meets the requirements of the college, a critical score has been determined which is exceeded by 50 per cent of the group. Those students whose scores on the test place them below the critical score are required to enroll in the basic course in mathematics.

Because test scores are not infallible, some individuals may be incorrectly exempted, while others may be incorrectly required to take the course. From the student's point of view such errors

---

[1] An outline of these skills is presented in the Appendix, pp. 126 ff.

[2] The Basic Skills in Mathematics Test, Form D, is presented in the Appendix, pp. 140 ff.

[Example of Types of Headings]

CHAPTER III

CHAPTER TITLE

<u>Center Heading</u>

This page represents the recommended form for a standard research report. The numerical designation, CHAPTER III, is centered in capital letters two inches from the top of the page. The chapter title is centered in capital letters two spaces below the numerical designation. The center heading is three spaces below the chapter title and is underscored. Three spaces are allowed between the center heading and the first line of context. Paragraphs are indented five spaces.

<u>Marginal Heading</u>

Marginal headings are placed flush with the left-hand margin two spaces below the last line preceding and are underscored. Two spaces are allowed between the marginal heading and the first line of the next paragraph.

<u>Center Heading</u>

<u>Marginal Heading</u>

<u>Paragraph heading.</u> A paragraph heading is placed two spaces below the marginal heading of the preceding paragraph, indented as for a paragraph, and underscored, but the words are not capitalized. The paragraph material continues immediately after the heading as shown.

[Example of a Table]

Table 1

Frequency of Scores, Means, and Standard Deviations of the Fourteen "Original" Items Included in the Tryout Administration of the Revised-Items Test

Scores	High School	College
0	1	0
1	1	1
2	3	6
3	10	6
4	14	21
5	26	34
6	32	45
7	33	59
8	43	61
9	28	48
10	32	41
11	21	20
12	8	17
13	7	3
14	3	9
Total	262	371
Mean	7.72	7.78
S.D.	2.64	2.51

[This table provides the numerical data for the figure shown on the following page. When inserted in a thesis it would be desirable to have it typed on the opposite side of the page facing the figure. However, if the table should be included with intervening pages of context materials, it should be placed as illustrated.]

[Example of a Figure]

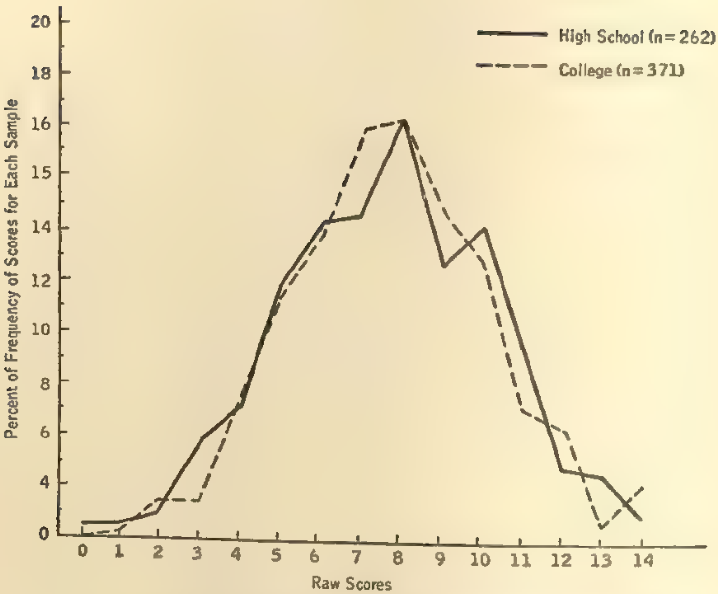


Figure 1. Comparison of Score Distributions for High School and College Samples.

[Example of a Figure]

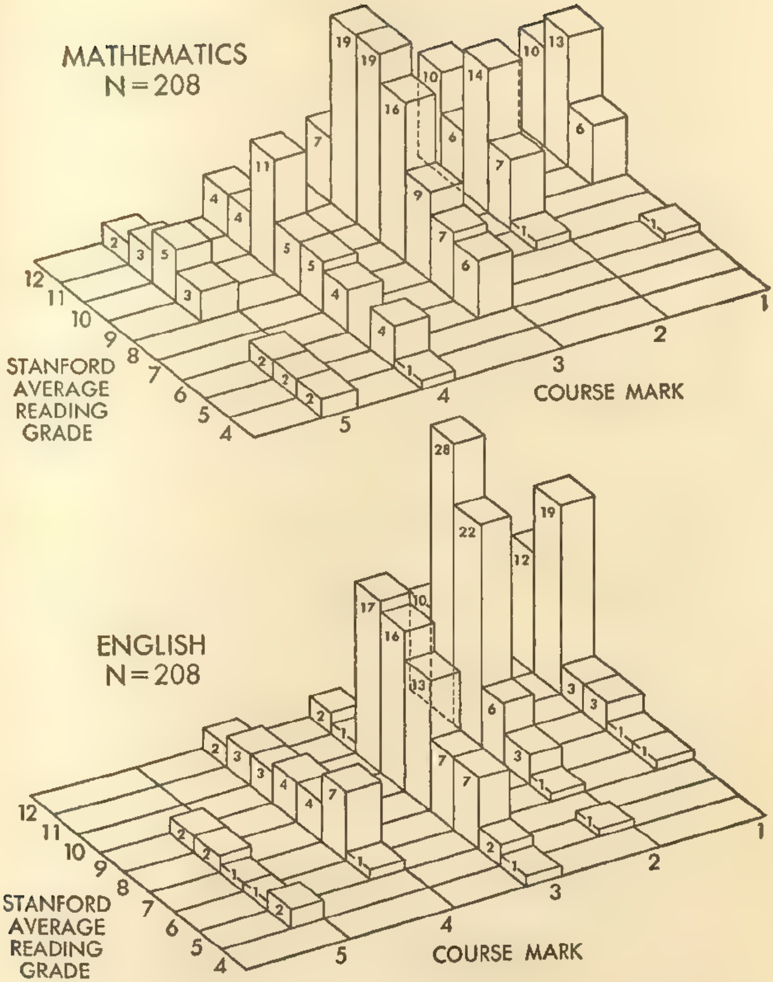


Figure 2. Distribution of Reading Grade Placement Scores on the Stanford Achievement Test in Ninth-Grade Mathematics and English Classes.

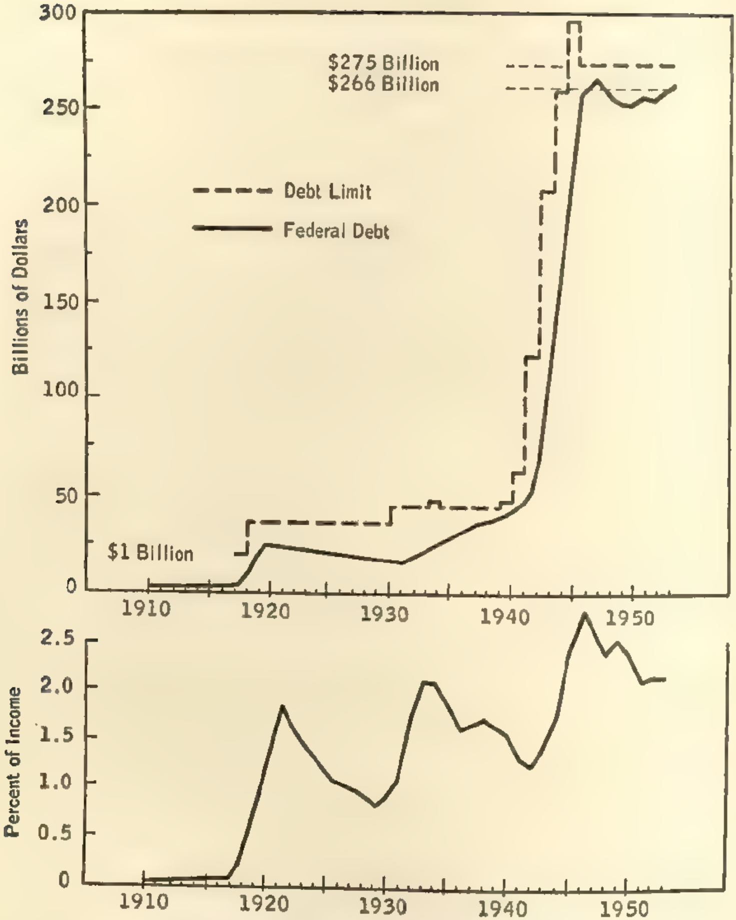


Figure 3. United States Federal Debt, Debt Limit, and Interest on Debt as Percentage of National Income: 1910-1953.
(Sources: Treasury Department; Department of Commerce.)

[Example of a First Page of a Bibliography]

BIBLIOGRAPHY

Brogden, Hubert E. "A Rationale for Minimizing Distortion in Personality Questionnaire Keys." Psychometrika, 19:141-148, 1954.

Garrett, Harry E. "The Discriminant Function and Its Use in Psychology." Psychometrika, 8:65-79, 1953.

Horst, Paul. "Mathematical Contributions." The Prediction of Personal Adjustment. Social Science Research Council Bulletin No. 48. New York: the Council, 1941. Pp. 403-447.

Kelley, E. Lowell and Fiske, Donald D. "The Prediction of Success in the VA Training Program in Clinical Psychology." American Psychologist, 5:395-406, 1950.

Lindquist, Everet F. Educational Measurement. Washington, D.C.: American Council on Education, 1951. xx + 819 pp.

[Other items of bibliographical reference were included in the same manner as above.]

[WHOLE BOOK CITED]

[Single Author, Named]

Stroud, James B. Psychology in Education. New York: Longmans, Green & Co., Inc., 1946. vii + 664 pp.

[Two or Three Authors, Named]

Jacobson, Paul B.; Reavis, William C.; and Logsdon, James B. The Effective School Principal. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1954. xxii + 617 pp.

[More than Three Authors, Only First Named]

Koos, Leonard V. et al. [or, and others.] Administering the Secondary School. New York: American Book Company, 1940. xii + 678 pp.

[Author as Editor]

Lindquist, Everet E. (Editor). Educational Measurement. Washington, D.C.: American Council on Education, 1951. xix + 819 pp.

[Author and Editor]

Adler, Alfred. What Life Should Mean to You. Edited by Alan Porter. Boston: Little, Brown & Co., 1931. 300 pp.

[Author a Committee or Association]

National Education Association, Education Policies Commission. The Purposes of Education in American Democracy. Washington, D.C.: The Association, 1938. 154 pp.

[Anonymous Author]

Out of Their Own Mouths: Utterances of German Rulers, Statesmen, Savants, Publicists, Journalists, Poets, Businessmen, Party Leaders, and

Soldiers. New York: Appleton-Century-Crofts, Inc.,1917. xxii + 254 pp.

[Supplementary Information Necessary]

Roget, Peter Mark. Thesaurus of Words and Phrases. Enlarged by John Lewis Roget: New edition revised and enlarged by Samuel Romilly Roget. Revised and authorized American edition. New York: Grosset & Dunlap, 1947. xxix + 705 pp.

[PART OF BOOK CITED]

[Title of Part Given]

Kambly, Paul E. "Marking and Reporting." The American Secondary School. Edited by Paul B. Jacobson. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1952. Pp. 373-394.

[Title of Part Not Given]

Good, Carter V. and Scates, Douglas E. Methods of Research, pp. 866-874. New York: Appleton-Century-Crofts, Inc., 1954. xx + 920 pp.

[BULLETIN OR YEARBOOK]

[Whole Cited, Author Named]

Long, John A. Conducting and Reporting Research in Education. Bulletin No. 6 of the Department of Educational Research. Toronto, Canada: University of Toronto, 1936. v + 77 pp.

[Whole Cited, Author Not Named]

1952 Invitational Conference on Testing Problems. Proceedings. Princeton, N.J.: Educational Testing Service, 1953. 138 pp.

A Guidance Aid for High Schools in Oregon. Salem, Oregon: State Board of Education, State Division of Vocational Education, March 1948. 38 pp.

[Part Cited, Author of Part Named]

Rummel, J. Francis. "Streamlining State Dental

Board Examinations." Proceedings of the Seventieth Annual Meeting of the American Association of Dental Examiners. Cleveland, Ohio: The Association (Dr. Carl A. Bumstead, Lincoln, Nebraska, Secretary), 1953.

[ENCYCLOPEDIA ARTICLE]

[Author Named]

"Science of Education," by B. Othanel Smith. Encyclopedia of Educational Research, Revised Edition, pp. 1145-1152. Edited by Walter S. Monroe. New York: The Macmillan Company, 1950.

[Author Not Named]

"The Renaissance." Encyclopaedia Britannica, 1952 Edition, Vol. 19, pp. 122-135. Chicago: Encyclopaedia Britannica, Inc.

[PERIODICAL REFERENCE]

[Author Named]

Brogden, Hubert E. "A Rationale for Minimizing Distortion in Personality Questionnaire Keys." Psychometrika, 19:141-148, June 1954.

[Author Not Named]

"Russia Training More Scientists." The Phi Delta Kappan, 36:180, February 1955.

[EDITORIAL OR NEWS ITEM]

[Author Named or Credited]

"Treaty With Chiang Raises Questions of Involvement," by Marquis Childs. The Oregonian, February 8, 1955, Sect. 1, p. 12. (Published in Portland, Oregon.)

[Author Not Named or Credited]

"Back to Mediocrity." Editorial. The Oregonian, February 8, Sect. 1, p. 12. (Published in Portland, Oregon.)

[THESES AND DISSERTATIONS]

[Published]

Sobel, Frances S. Teachers' Marks and Objective Tests as Indices of School Adjustment. Contributions to Education, No. 674. New York: Bureau of Publications, Teachers College, Columbia University, 1936. viii + 74 pp.

[Unpublished]

Carvey, Sister M. Virginia Claire. A Survey of Problems of Parents of Mentally Deficient Children and of Certain Services to These Parents. Unpublished Doctor's Dissertation. Eugene, Oregon: The University of Oregon, 1953. 223 pp. Typewritten.

Connett, Irene. An Experimental Study on Reading Readiness of the First Graders in Eugene, Oregon. Unpublished Master's Field Study. Eugene, Oregon: The University of Oregon, 1952. 29 pp. Typewritten.

[MISCELLANEOUS]

[Pamphlet]

Huxley, Julian; Faust, Clarence H.; and Chaffee, Zechariah. Charter Day Addresses. Delivered at the University of Oregon, October 18-19, 1954. Eugene, Oregon: University Editor, The University of Oregon. 36 pp.

Zimand, Gertrude Folks. Young Workers in the United States. Annual Report of the National Child Labor Committee for the year ending September 30, 1953. Publication No. 412, November 1953. 20 pp. New York: National Child Labor Committee. (419 Fourth Avenue, New York 16, N.Y.)

"Methods of Expressing Test Scores." Test Service Bulletin, No. 48, January 1955. New York: The Psychological Corporation. 6 pp.

[Test]

SRA Achievement Series. (Arithmetic, Language Arts, Reading, Work-Study Skills, Grades 4-6, 7-9). Chicago: Science Research Associates, 1954.

Read General Science Test. Written by John G. Read. Evaluation and Adjustment Series. High School level. Yonkers-on-Hudson, New York: World Book Company, 1954.

[Film]

You and Your Mental Abilities. 35mm. filmstrip. To accompany SRA Primary Mental Ability Tests. Chicago: Science Research Associates. (57 W. Grand, Chicago, Illinois.)

How to Read a Book. 16mm. film, sound, color, 10 minutes. Senior high school and adult level.

Collaborator: Dr. William Brink, Professor of Education, Northwestern University. Chicago: Coronet Productions, 1947. (65 E. South Water Street, Chicago, Illinois.)

[Chart]

Diagnostic Profile Chart for California Achievement Tests. Elementary form, Grades 4-6. 21½ by 30 inches, color. Los Angeles: California Test Bureau, 1950.

APPENDIX B

Review of Basic Statistical Concepts and Computation

Frequency Distribution

- Mechanics for constructing frequency distributions
- Histogram (bar graph)
- Frequency polygon (line graph)
- Ogive or cumulative frequency curve
- Characteristics of common frequency distributions

Measures of Central Tendency

- The mode
- The mean
- The median

Percentiles, Deciles, and Quartiles

- Computational procedures

Measures of Variability

- Range (R)
- Interquartile range
- Semi-Interquartile range (Q)
- Average deviation (AD or MD)
- Standard deviation (SD or σ)
- Relationship between Q and σ

Standard and Scaled Scores

- *z*-Scores
- T-Scores
- Stanines

Measures of Relationship

Computation of r

Computation of rho (ρ)

Use of correlation coefficients in prediction

Cautions in the use and interpretation of correlation coefficients

Testing Significances of Differences

The *t*-test

Level of significance

Degrees of freedom

In research work some basic statistical concepts are used a great deal because they are helpful in summarizing the information and data obtained. This appendix makes no pretense of being a complete treatment of all the statistical tools necessary in research work, but rather presents some of the more basic concepts as an introduction (or review) prior to the more detailed and extensive study on the part of the researcher. The computational procedures presented here are only for the purpose of illustrating the statistical concepts. The researcher should refer to standard textbooks on statistical methodology for more extensive and detailed instruction dealing with various statistical procedures.

FREQUENCY DISTRIBUTION

In most research studies, considerable data are assembled concerning some quantitatively expressed variables. As these data are collected they are usually in such a haphazard order that they provide only the vaguest idea of numerical size and order. It is difficult for a researcher to derive any adequate idea of the performance of a group as a whole from the unorganized form in which the research data are collected. One method for organizing quantitative data consists of the preparation of a frequency distribution which may be presented either in tabular or graphical form.

While frequency distributions can be constructed for many types of data, assume for purposes of illustration that a given set of data consists of the scores made by 80 pupils on a test of 50 words in spelling, as shown in Table 5. It is obvious from

TABLE 5. Scores of 80 Fourth-Grade Pupils on a 50-Word Spelling Test

22	25	16	44	43	6	42	39
49	46	39	37	35	32	29	26
17	22	28	30	32	35	27	38
45	41	16	21	25	38	37	35
31	30	28	40	38	36	13	24
27	30	31	35	31	31	30	30
40	37	36	34	20	24	27	30
18	22	36	29	37	36	33	11
45	40	19	23	27	29	36	34
19	23	27	29	39	37	36	33

this unorganized set of scores that it would be difficult to determine what score would be most representative of the group as a whole or to determine the amount of variability among the scores. Table 6 shows three different sets of the same data in which the scores have been grouped by two's, by three's, and by five's.

In the first group (by two's) it appears that there is a concentration of pupils having scores of 30 and 31, and another concentration having scores of 36 and 37, with the other scores scattered out fairly evenly except at the extremes. In the second group (by three's) the concentrations about scores of 30 and 36 are shown as before, with more tapering off toward the extremes. In the third group by five's) there appears to be only one concentration with a tapering off toward the extremes. In the latter case the characteristic of the group having two concentrations is obliterated by the grouping process. In the first set, the general characteristics of the group do not appear as sharply as in the other two sets. It has usually been believed by statisticians and researchers that there is no real need for more than 20 categories, or intervals, of data and

TABLE 6. Grouped Frequency of Spelling Scores in Table 5

Intervals of 2			Intervals of 3			Intervals of 5		
x	Tabulation	f	x	Tabulation	f	x	Tabulation	f
48-49	/	1	47-49	/	1	48-52	/	1
46-47	/	1	44-46	////	4	43-47	~~////~~	5
44-45	///	3	41-43	///	3	38-42	~~////~~ ~~////~~ /	11
42-43	//	2	38-40	~~////~~ ////	9	33-37	~~////~~ ~~////~~ ~~////~~ ////	19
40-41	////	4	35-37	~~////~~ ~~////~~ ~~////~~	15	28-32	~~////~~ ~~////~~ ~~////~~ ///	18
38-39	~~////~~ /	6	32-34	~~////~~ /	6	23-27	~~////~~ ~~////~~ //	12
36-37	~~////~~ ~~////~~ /	11	29-31	~~////~~ ~~////~~ ////	14	18-22	~~////~~ ///	8
34-35	~~////~~ /	6	26-28	~~////~~ ///	8	13-17	////	4
32-33	////	4	23-25	~~////~~ /	6	8-12	/	1
30-31	~~////~~ ~~////~~	10	20-22	~~////~~	5	3- 7	/	1
28-29	~~////~~ /	6	17-19	////	4			80
26-27	~~////~~ /	6	14-16	//	2			
24-25	////	4	11-13	//	2			
22-23	~~////~~	5	8-10		0			
20-21	//	2	5- 7	/	1			
18-19	///	3			80			
16-17	///	3						
14-15		0						
12-13	/	1						
10-11	/	1						
8- 9		0						
6- 7	/	1						
		80						

that many important group characteristics may be hidden by the use of less than 10 intervals.

Mechanics for Constructing Frequency Distributions

The following steps are usually followed in the construction of frequency distributions, although there are some variations and exceptions in the case of "natural" groupings and some types of data. For detailed instructions in handling exceptional cases, refer to the indexes of standard textbooks on statistical methodology.

1. Set up a data sheet as shown in Table 6 with column headings of *X* (for scores), Tabulation (for tally marks), and *f* (for the frequency of the scores in each interval).
2. Determine the range of scores. For this purpose the range is the numerical difference between the highest and lowest scores. (In the example in Table 5, the range would be the difference between 49 and 6, or 43.)
3. Determine the size of the interval that is to be used for grouping the measures to simplify computational procedures and to show the basic characteristics of the data. The range is commonly divided by 15 to determine the approximate size of the intervals to use. (By dividing 43 by 15, the size of the interval in our example would be approximately 3.) The actual size of the interval we would use, however, would be the *preferred interval* closest to the size we obtained by dividing our range by 15. *Preferred intervals* for various sets of data are *odd* numbers such as 1, 3, 5, and 7.
4. Determine the midpoints of these intervals so that they are multiples of the interval size. (In this case, we would use an interval of 3, and the midpoints of all intervals would be multiples of 3.)
5. Write the integral value limits for each interval, in descending order, in the column headed *X* on the data sheet. (The integral limits would be, in this case, the scores immediately above and below the midpoints or multiples of 3 as shown in the central distribution in Table 6.) It is also customary to indicate these intervals only by the value of the midpoint instead of the limits of the interval. This makes a table less crowded, but one must remember that the midpoint value is representing the entire interval of several scores.

6. Tally the scores from the unorganized list by placing a tally mark in the column headed Tabulation and opposite the score limits in which each given score would fall. (For example, the first score, 22, would be tallied at the right of 20–22 as shown in Table 6. Proceed accordingly for the remaining values in the original list.)
7. Count the tally marks opposite each interval and write the result in the column headed *f*. The total of all frequencies in this column should equal the total number of scores in the original list.

Histogram (Bar Graph)

A histogram, or bar graph, is a graphical representation of a frequency distribution. The histogram illustrated in Figure 32 is based on the middle frequency distribution given in Table 6. These may be made on graph paper for the analysis of research data, but should be shown in the final written report of the research with only the midpoints of the intervals shown on the base line and the frequencies indicated by a vertical scale at the left of the graph. The histogram is constructed by first laying off the midpoints of the intervals at equal distances along a base line. The value of the midpoint of the interval above the highest interval for which there is a frequency indicated and the midpoint of the interval below the lowest interval for which a frequency is indicated should be included. Therefore, the midpoints for intervals of scores 50–52 and 2–4 are included on the base line as shown in Figure 32. The vertical scale is erected immediately above the extreme midpoint at the left-hand end of the base line. The vertical scale, indicating the frequencies of the intervals, should start with zero frequency at the base line and extend upward to include the largest class or interval frequency. From the diagram it is easy to note general characteristics of the data.

Frequency Polygon (Line Graph)

Another commonly used graphical description of the frequency distribution is the frequency polygon, or line graph. The one shown in Figure 33 is based on the same data as the histogram in Figure 32. Instead of drawing vertical bars to

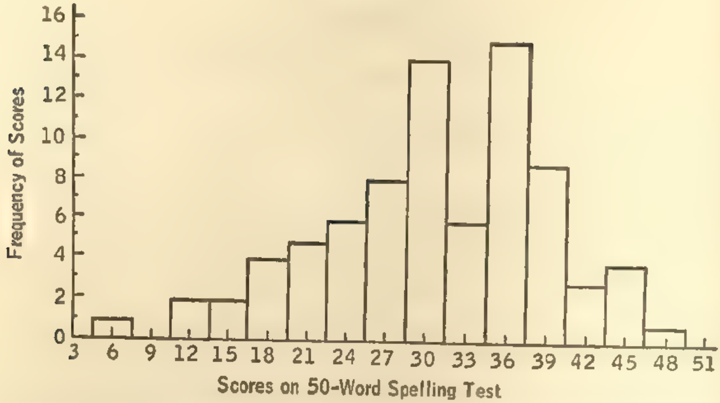


FIGURE 32. Histogram of Scores of 80 Pupils on a 50-Word Spelling Test.

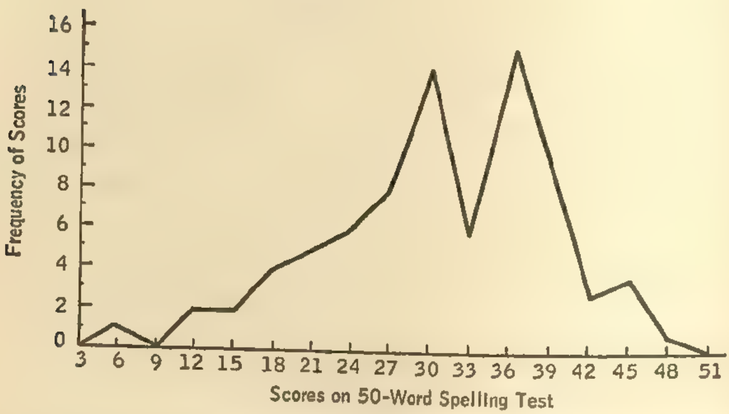


FIGURE 33. Frequency Polygon of Scores of 80 Pupils on a 50-Word Spelling Test.

indicate the frequencies of the intervals, a line is drawn connecting the points representing the midpoints of the upper bases of the bars. These points are placed at the intersections of lines that would extend horizontally to the right of the frequency scale and vertically from the midpoints of the interval scale. The polygon is *closed* at each end by extending the graph line to the base line at the midpoint of each of the

extreme intervals with zero frequencies immediately adjacent to the intervals containing the given frequencies.

Ogive or Cumulative Frequency Curve

A less used but extremely valuable graphical representation of data for some purposes is the ogive. This curve is often used in determining percentile ranks corresponding to given scores or interval midpoints and is sometimes referred to as the percentile curve. In preparing for construction of this curve it is necessary to add another column to the frequency distribution table to record the cumulative frequencies in ascending order. The frequency for the lowest interval is recorded in the *cf* (cumulative frequency) column. The frequency for the next higher interval is added to the cumulative frequency of the interval below. This process is carried on to the highest interval, whose *cf* should then equal the total number of values in the distribution. The curve is plotted in much the same way as that for the frequency polygon, except that the points joined by the straight lines are plotted directly above the *upper real limit of each interval instead of its midpoint.*

At this point it is desirable to determine the *real limits* of the intervals as distinguished from the integral limits. The real limits are usually considered to be halfway between the lower integral limit of one interval and the higher integral limit of the next lower interval. Thus the real limits of the interval 32–33–34, whose midpoint is 33, are 31.5 and 34.5, and the real limits of the next higher interval, whose midpoint is 36, are 34.5 and 37.5. The frequency distribution, cumulative frequency, and ogive for the data presented in the preceding figures are shown in Figure 34.

Since the ogive is also referred to as the percentile curve, it is possible to determine percentile points for various values, or the values having a given percentile rank, by dividing the vertical axis or frequency scale into 100 equal parts. In Figure 34, this has been done at the right-hand side of the figure and labeled as the cumulative percentage of frequencies. To obtain the value having any given percentile rank it is necessary only

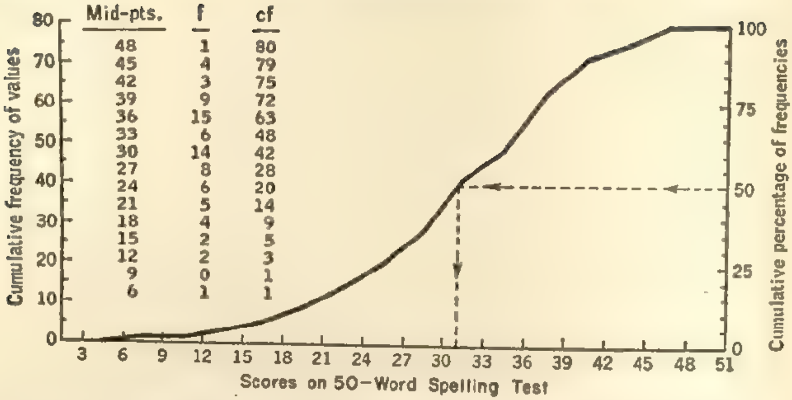


FIGURE 34. Cumulative Frequency Distribution and Ogive of Scores of 80 Pupils on 50-Word Spelling Test. The dashed lines locate the median score or 50th percentile.

to move horizontally from the given cumulative percentage to the left to the ogive curve and then down to the score scale. The dotted line on the graph illustrates this process for determining the 50th percentile or median of the distribution.

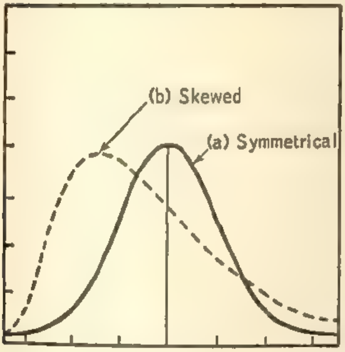
Characteristics of Common Frequency Distributions

The types of data found in research commonly show a tendency to group about a given point in a distribution. This grouping tendency gives rise to a peak which usually occurs in a graphical representation of a frequency distribution and makes possible the use of a "typical" value to describe a mass of data, or a group as a whole. The location of this point in a distribution may be identified as an "average." Some of the more common characteristics of frequency distributions are indicated by the following terms which are illustrated by Figures 35 to 42: symmetrical, skewed, bell-shaped, normal, rectangular, U-shaped, and J-shaped distributions, and the characteristics of dispersion, skewness, and kurtosis.

Symmetrical distributions are those in which the frequencies of corresponding intervals on opposite sides of a middle line are equal, as shown in Figure 35, curve *a*.

Skewed distributions are those in which the extreme values

of the distribution extend in one direction more than in the other. If the extreme values extend more to the right than to the left, the distribution is described as being *positively* skewed or *skewed to the right,* as shown in Figure 35, curve *b.* The extension of extreme values to the left forms a *negatively* skewed or *skewed to the left* type of distribution.



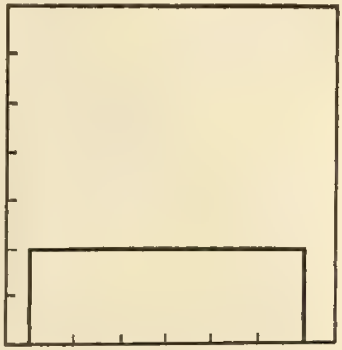


FIGURE 35. (*Left*) A Symmetrical and Positively Skewed Distribution.
FIGURE 36. (*Right*) Rectangular Distribution.

There is a large family of *bell-shaped* distributions, of which the normal distribution, or *normal curve,* has received considerable attention, has been extremely useful as a basis for the analysis of certain kinds of data, and has been misused excessively. Figure 35, curve *a,* illustrates a normal curve. This is a precise curve based on a specific mathematical relationship between the values on the horizontal and vertical axes and described by the following equation:

$$Y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

Y is the height of the curve corresponding to an assigned value on the horizontal axis

π is a constant approximately equal to 3.1416

e is the base for "natural" logarithms, a constant approximately equal to 2.7183

σ is the standard deviation of the distribution

z is the ratio of the difference between an assigned value (X) on the horizontal axis and the mean (M) of the distribution to the standard deviation, or

$$z = \frac{X - M}{\sigma}$$

Rectangular, U-shaped, and *J-shaped* distributions are illustrated in Figures 36, 37, and 38 respectively, and are self-evident.

Dispersion refers to the degree of variation in size of the values in a distribution. As seen in Figure 39, the values of the variable included under curve *a* vary to a greater degree than the variable included in curve *b*. That is, there is a greater spread, or scatter, to the sizes of the values in curve *a* than in curve *b*.

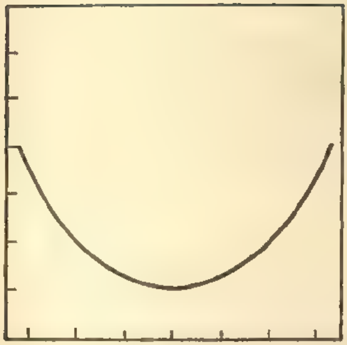
Skewness refers to the degree of nonsymmetry of a curve as described above, and illustrated by Figure 35.

Kurtosis indicates whether a distribution is relatively peaked or flat-topped. A *normal curve* is called *mesokurtic,* which means that it is neither very peaked nor very flat across the top. Curves that are more flat than the normal curve are called *platykurtic,* and those that are more peaked are called *leptokurtic.* These are shown in Figure 40.

MEASURES OF CENTRAL TENDENCY

When it is desired to describe some general characteristic of a group of children (distribution of data) it may be noted immediately that there are differences among individuals. For example, a test of reading given to a class of pupils in the sixth grade of a school would reveal that some pupils were quite good readers and others were quite poor readers. It is obvious that not all pupils would obtain the same score on a test of reading.

One way of describing the reading proficiency of the *group* would be to list the score of each individual child. If the group is large, this is a cumbersome method, difficult to interpret, and not very useful for many purposes. A better description might be the "average" score for the group. However, in statistical



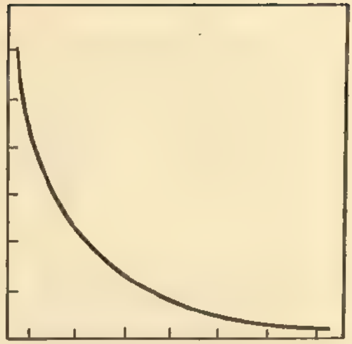
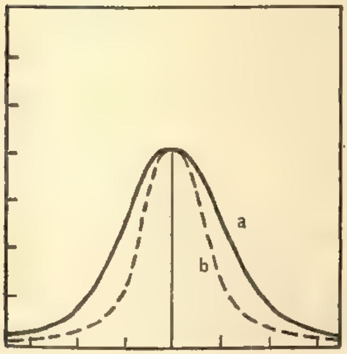


FIGURE 37. (*Left*) U-Shaped Distribution.
FIGURE 38. (*Right*) J-Shaped Distribution.



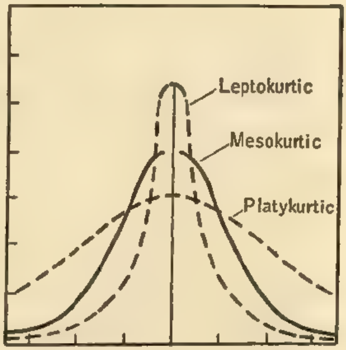
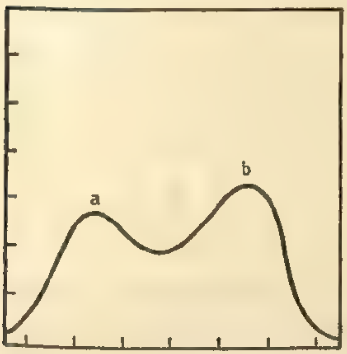


FIGURE 39. (*Left*) Illustration of Dispersion.
FIGURE 40. (*Right*) Illustration of Kurtosis.



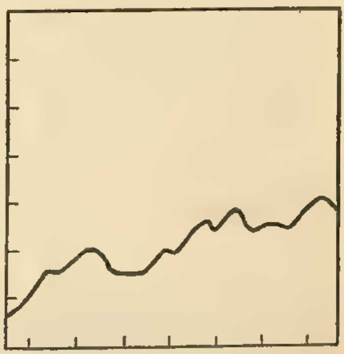


FIGURE 41. (*Left*) Bimodal Distribution.
FIGURE 42. (*Right*) Irregular Distribution.

language, any measure that gives a value representative of a group as a whole, or of things in general, is called an average. Three measures of average, or cental tendency of a group of values, are the mode, the mean, and the median.

The Mode

The mode is defined as that value in a distribution for which there is the largest frequency. In multimodal distributions the two or more values having the higher frequencies indicate the different modes. The mode is easily found by inspection of a frequency distribution tally chart, or by simply counting to see which value occurs most often. It is a useful measure when only a quick, preliminary measure is desired, and it provides a little-better-than-guess estimate of the group as a whole. It cannot be used for further statistical work, and it is quite unreliable as a measure for research work.

The Mean

A better description of the group may be obtained by making use of all values in the distribution. The common average, obtained by adding together all values in the distribution and dividing by the number of them, is more precisely known as the *arithmetical mean*. This measure is usually indicated by the symbols, M or $\overline{X}$.

For purposes of illustration in an example requiring very little computational effort, suppose that a group consists of 12 individuals as shown in Group I, below. Suppose, also, that scores are obtained for each of them on some measurement as follows:

Group I

Individuals:	A	B	C	D	E	F	G	H	I	J	K	L
Scores:	10	20	30	40	50	60	70	80	90	100	110	120

By adding all the scores (total 780), and dividing by the number of individuals (12), the mean is found to be 65. When a group consists of a large number of individuals, the mean is sometimes found by a "grouping" method as follows (refer to

TABLE 7. Deviation Method for Computation of the Mean and the Standard Deviation

X Mid Pts.	f	d	fd	fd^2	cf[a]
48	1	6	6	36	80
45	4	5	20	100	79
42	3	4	12	48	75
39	9	3	27	81	72
36	15	2	30	60	63
33	6	1	6	6	48
30(AR)	14	0	0	0	42
27	8	−1	−8	8	28
24	6	−2	−12	24	20
21	5	−3	−15	45	14
18	4	−4	−16	64	9
15	2	−5	−10	50	5
12	2	−6	−12	72	3
9	0	−7	0	0	1
6	1	−8	−8	64	1
	80		20	658	
	N		Σfd	Σfd^2	

$$M = AR + i\left(\frac{\Sigma fd}{N}\right) \qquad \sigma = i\sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$$

$$M = 30 + 3\left(\frac{20}{80}\right) \qquad \sigma = 3\sqrt{\frac{658}{80} - \left(\frac{20}{80}\right)^2}$$

$$M = 30.75 \qquad \sigma = 8.58^{-}$$

[a] Used for computation of *Mdn* and other percentiles.

Table 7 for an illustration of this procedure with the values given in Tables 5 and 6, previously):

1. Arrange the values in a grouped frequency distribution according to instructions given on pp. 324–325.
2. Select some arbitrary reference point (AR) as the midpoint of an interval of values as near the center of the distribution as possible. The interval with the greatest frequency is often the best one to pick.

3. Mark in the deviation (d) column the number of intervals each interval deviates from the interval whose midpoint has been selected for the arbitrary reference point. The value of the AR interval is marked zero in this column. Values above AR are marked plus and those below are marked minus.
4. Multiply each interval deviation (d) by its corresponding frequency (f) and write the product in the next column (fd).
5. Find the algebraic sum of the plus and minus fds.
6. Divide the result of step 5 by the number of scores (N) in the total distribution. This gives the correction $\left(\frac{\Sigma fd}{N}\right)$ in *interval units* of the mean of the distribution from the arbitrary reference point.
7. Multiply the correction value obtained in step 6 by the number size of the intervals (i). This gives the correction in *value units* of the mean from arbitrary reference point.
8. Add the result of step 7 algebraically to AR. The result is the *arithmetical mean* of the distribution of values.

The formula corresponding to these steps is as follows:

$$\text{Mean } (M) = AR + i\left(\frac{\Sigma fd}{N}\right)$$

The Median

Another measure of central tendency is that point in a distribution of values above and below which half the cases fall. Since the values in Group I are already arranged in order of their magnitude, it is readily seen that this point would fall between values of 60 and 70 for individuals F and G, respectively. It is common procedure to use the value halfway between, or 65 in this case, to represent the *median* of the distribution. This term is usually indicated by the symbol, Mdn. It should be noted in this particular illustration that the distribution is symmetrical (also rectangular) and, therefore, the mean and the median have the same value.

When distributions are not perfectly symmetrical the mean and the median have different values, and a problem arises as to which value is the better to describe the distribution or to use in further statistical calculations. The choice of using the mean or the median depends to a great extent upon the shape

of the distribution and the distortion effects of extreme scores. The choice also depends upon the need for further arithmetical processing, such as combining averages or making statistical computations requiring a measure of central tendency. The median cannot be used for these purposes.

Consider, for example, another group of 12 individuals with values as shown in Group II.

Group II

Individuals:	A	B	C	D	E	F	G	H	I	J	K	L
Scores:	10	12	12	14	14	14	16	23	28	30	91	96

The sum of these twelve values is 360, yielding a *mean* equal to 30. The *median* falls between the values for individuals F and G, or values 14 and 16, and is equal to 15. In this particular illustration, the mean value is twice as large as the median so that it is obvious that one measure cannot be as representative of the group as a whole as the other. If the mean were to be used to describe the group as a whole, it would be noted that only two individuals have values above the mean and nine have values below the mean. It is also apparent that individuals K and L would be considered exceptional in whatever was measured in comparison with the rest of the group. Aside from these two individuals, the values all fall between 10 and 30, so that some value between 10 and 30 would probably be a more representative measure of the group than any measure higher than this range of values. In this particular case the median would be the more representative measure. As a general rule, however, the mean is the more desirable measure of central tendency, except when there are a few extreme values that are not representative of the group.

PERCENTILES, DECILES, AND QUARTILES

It is often desirable to be able to describe a distribution at some point or points other than those given for measures of central tendency. There are occasions when it may be desirable to know the point below which any given percentage of cases lie; 25 percent, for example. A number of measures have been developed to describe a distribution at practically any place

along the scale of values, the more frequently used being percentiles, deciles, and quartiles.

The *percentile,* or centile, is defined as that point in a distribution below which a given percent of cases lie. For example, the tenth percentile is that point in the distribution of values below which lie 10 percent of the cases. Accordingly, the median of a distribution may also be referred to as the fiftieth percentile. When the rank of any given value is based on the division of all values into 100 consecutive groups, that rank is called the *percentile rank* of the value.

When values are divided into 100 consecutive groups, and then subdivided into 10 equal groups, each of the subgroups is separated from its adjacent groups by values called *deciles.* The first decile is that point in the distribution of 10 subgroups below which lie 10 percent of the cases; the second decile is that point below which 20 percent of the cases lie; and so on.

When these values are subdivided into four consecutive groups, each group is separated from its adjacent group by values referred to as *quartiles.* The first quartile is that point in a distribution below which 25 percent of the cases lie, and it is denoted by the symbol, Q_1. The second quartile is that point below which 50 percent of the cases lie, and it is the *median,* denoted by either the symbol *Mdn,* or Q_2. The third quartile is that point below which 75 percent of the cases lie, and it is symbolized by Q_3

Computational Procedures

The simplest method of computing percentiles is to place all values in descending order and then to compute the proportion of the total number that falls below any given value. The *percentile rank* for any given value, then, is the proportion thus obtained.

In calculating the value for any given percentile rank from a grouped frequency distribution, the following formula may be used according to the steps listed below.

$$V = LL + i\left(\frac{pN - lcf}{f}\right)$$

V is the value for any given percentile rank desired

pN is the proportion of the total number of cases lying below the desired percentile point. (For example, to find Q_3 we would want to find the value below which .75N lie.)

LL is the lower real limit of the interval in which the given percentile rank lies

lcf is the cumulative frequency of the interval *below* the one in which the given percentile rank lies

f is the frequency of values in which the given percentile rank lies

i is the number of values in an interval

To obtain the value corresponding to Q_3, or the 75th percentile rank, the following steps may be used. These may also be applied in finding the value for any other percentile rank by using a different value of pN in step 1.

1. Determine the number of cases (pN) corresponding to the desired proportion of the total number for the given percentile rank. (To find Q_3 for the distribution shown in Table 7, determine .75 of the total number [80], which would be 60.)
2. Find the interval in the cf column which would contain the pN^{th} case. (This would be the interval whose cf is 63 and whose midpoint is 36.)
3. Record the value of the lower real limit (LL) of the interval in step 2. (Since this interval contains scores 35, 36, and 37, the lower real limit would be 34.5.)
4. Substract the partial cf, or lcf, (up to but not including the interval in step 2) from pN. (This would be $60 - 48$, or 12.)
5. Divide the result of step 4 by the f in the interval containing the given percentile rank. (Thus, 12/15 or .8.)
6. Multiply the result of step 5 by the size of the interval, i. (Thus, .8 x 3 or 2.4.)
7. Add the result of step 6 to the result of step 3, the lower real limit of the interval containing the given percentile rank. (This would be $34.5 + 2.4$ or 36.9, the value of Q_3.)

As represented by the formula, this would be:

$$Q_3 = 34.5 + 3\left(\frac{60 - 48}{15}\right) = 36.9$$

MEASURES OF VARIABILITY

Measures of central tendency do not reveal anything at all about the amount of spread, scatter, dispersion, or variability of a group of values. Two groups may have the same measure of central tendency but be quite different in variability. One group may be homogeneous with respect to the characteristic under investigation, while another group may be quite heterogeneous. These differences are very important when it comes to analyzing the data for research studies. Consider a third group of 12 individuals and their corresponding values, as indicated below.

Group III

Individuals:	A	B	C	D	E	F	G	H	I	J	K	L
Scores:	45	50	52	60	62	64	66	68	70	78	80	85

The mean and median values for this group are equal (65) and the same as obtained for Group I, shown previously on p. 264. It is clear, however, that Group I is not exactly like Group III. Therefore it is necessary to have some measure other than measures of central tendency to describe these differences.

RANGE (R)

The simplest method of describing the amount of variability within a group is to find the *range* of scores. The range is defined as the difference between the highest and lowest values.[1] In Group I, the range is 110, while in Group III it is 40. Thus, Group I would appear to have over twice as much spread in values as Group III. The range can provide a quick estimate of the variability of a group, but it tends to be unreliable and cannot be used for further arithmetical processing. If the extreme scores for Group III had been 10 and 120, instead of 45 and 85, the range for this group would have been equal to the range for Group I. But, with the exception of the extreme scores in both groups, Group III is still more homogeneous than Group I. Thus, it can be seen that the range is limited in its use.

[1] Sometimes the range is described as the difference between the highest and lowest values plus 1.

Interquartile Range

The range within which the middle half of the cases in a distribution fall is defined as the interquartile range; that is, the difference between Q_1 and Q_3. In Group I, it is seen that the middle six pupils have scores of 40, 50, 60, 70, 80, and 90. The values of Q_1 and Q_3 are 35 and 95, respectively. Thus, the interquartile range for Group I is 60. In Group III, the interquartile range is 74–56, or 18. Thus, for the middle half of the individuals in these two groups, the interquartile range indicated that the variability of Group I is between three and four times as great as that of Group III.

Semiinterquartile Range (Q)

The semiinterquartile range, or quartile deviation as it is often called, is one half the size of the interquartile range and denoted by the symbol, $Q.D.$, or more simply as Q. This measure of variability, like the range, is based upon only two values in a distribution. However, since the two determining values, Q_1 and Q_3, are much more stable than the extreme values determining the range, this measure is more adequate to use.

When the median is used as the measure of central tendency in describing a distribution of values, the semiinterquartile range (Q) should be used to describe the variability of that distribution. The basic reason for this general rule is that both measures are determined from a rank ordering of the values in the distribution and do not reflect the magnitude of those values.

Average Deviation (*AD* or *MD*)

The average deviation, or mean deviation as it is sometimes called, is the mean amount of the *absolute* deviations of all individual values in a distribution from the mean of the distribution. In determining the average deviation it is necessary to find out how far each value deviates from the mean of all scores. Thus, with Group III, for example, with a mean of 65, the highest score, 85, deviates 20 points. The sum of the absolute deviations of all scores for Group III totals 114 points and

yields an average deviation of 9.5, as shown in Table 8. The mean deviation is computed by the following formula:

$$MD = \frac{\Sigma|x|}{N}$$

Σ is the upper-case Greek letter, Sigma, which means in statistical language "the sum of"

x is the lower-case X which indicates the deviation of any score (X) from the mean (M), or ($x = X - M$)

N indicates the total number of cases in the distribution

When parallel bars are used with x, they indicate that the deviation is an absolute value, or that the direction of the deviation above or below the mean has been disregarded and that the algebraic signs have no significance in further arithmetical processing.

TABLE 8. Computation of Average Deviation

Individuals	Values (X)	Absolute Deviations from the Mean $\|x\|$
L	85	20
K	80	15
J	78	13
I	70	5
H	68	3
G	66	1
F	64	1
E	62	3
D	60	5
C	52	13
B	50	15
A	45	20
Total	780	114

Mean: $780/12 = 65$
MD: $114/12 = 9.5$

If the average deviation is to be used merely for describing the variability of a distribution, and no further statistical cal-

culations are to be made, a disregard of the algebraic signs for the deviations is not important. Since this measure is seldom computed except for an easily obtained estimate of variability, due to the above limitations, it is included here more for instructional value to the reader, who is likely to find it referred to in research literature, than as an essential measure of variability. In addition, it makes it easier to understand the concept of the standard deviation (given below) which is essentially a modification of the average deviation in which algebraic signs are taken into consideration.

Standard Deviation (*SD* or σ)

A more precise and more usable measure of variability is the standard deviation. It is defined as the "square root of the mean of the squares of the individual deviations from the mean of a distribution," and is indicated by the Greek lower-case letter, sigma or σ. This measure takes into account the size of every value in the distribution and is usable for further arithmetical processing and in many statistical calculations. In the computation of this measure the algebraic signs of the deviations must be taken into account. The computational formula is

$$\sigma = \sqrt{\frac{\Sigma x^2}{N}}$$

but the process of computing the standard deviation is simplified by use of the deviation method using the following formula and steps as listed below.

$$\sigma = i\sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$$

1. Set up a frequency distribution table, as given in Table 7, showing *f, d,* and *fd* columns.
2. Multiply each interval deviation (d) by the corresponding product of the frequencies and deviations (fd) and write the product in the next column headed fd^2.
3. Find the algebraic sum of the fd^2s.
4. Divide the result of step 3 by the number of values (N) in the total distribution.
5. Square the correction value found in step 6 in the computation of the *Mean,* as shown on p. 334. This gives $(\Sigma fd/N)^2$.

6. Subtract the value obtained in step 5 from the value obtained in step 4.
7. Extract the square root of the result of step 6.
8. Multiply the result of step 7 by the size of the intervals (i) to obtain the *standard deviation.*

Refer to Table 7 for an example of the computation of the standard deviation by this procedure.

When the mean is used as the measure of central tendency in describing a distribution of values, the standard deviation should be used to describe the variability of that distribution. The basic reason for this general rule is that both measures are determined by the magnitude of each and every value rather than the rank order of those values in the distribution. Note the analogous statement with respect to the median and Q on p. 339.

Relationship Between Q and σ

When the standard deviation of *normally* distributed data has been computed, there are some interesting and important relationships between values of certain size and the number of values in the distribution. Similarly, certain relationships exist between the number of values, the size of them, and the semi-interquartile range. These relationships are illustrated in Figure 43. The reader must be cautioned, however, that these

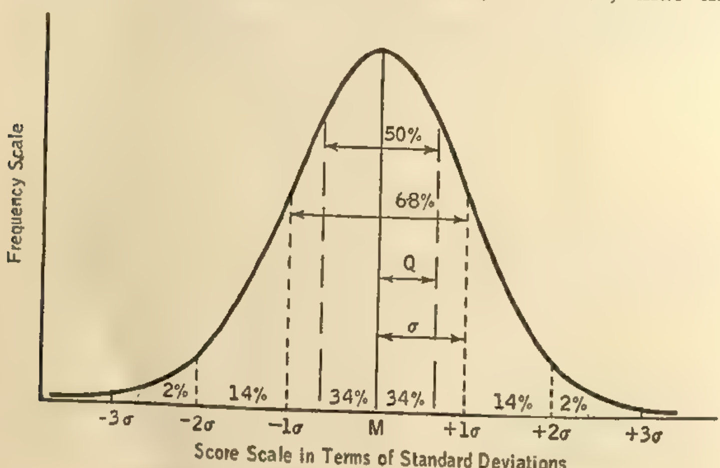


FIGURE 43. Normal Frequency Distribution.

TABLE 9. Percentage of Total Area Under the Normal Curve Between Mean Ordinate and Ordinate at any Given Sigma Distance from the Mean

$\frac{x}{\sigma}$	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	00.00	00.40	00.80	01.20	01.60	01.99	02.39	02.79	03.19	03.59
0.1	03.98	04.38	04.78	05.17	05.57	05.96	06.36	06.75	07.14	07.53
0.2	07.93	08.32	08.71	09.10	09.48	09.87	10.26	10.64	11.03	11.41
0.3	11.79	12.17	12.55	12.93	13.31	13.68	14.06	14.43	14.80	15.17
0.4	15.54	15.91	16.28	16.64	17.00	17.36	17.72	18.08	18.44	18.79
0.5	19.15	19.50	19.85	20.19	20.54	20.88	21.23	21.57	21.90	22.24
0.6	22.57	22.91	23.24	23.57	23.89	24.22	24.54	24.86	25.17	25.49
0.7	25.80	26.11	26.42	26.73	27.04	27.34	27.64	27.94	28.23	28.52
0.8	28.81	29.10	29.39	29.67	29.95	30.23	30.51	30.78	31.06	31.33
0.9	31.59	31.86	32.12	32.38	32.64	32.90	33.15	33.40	33.65	33.89
1.0	34.13	34.38	34.61	34.85	35.08	35.31	35.54	35.77	35.99	36.21
1.1	36.43	36.65	36.86	37.08	37.29	37.49	37.70	37.90	38.10	38.30
1.2	38.49	38.69	38.88	39.07	39.25	39.44	39.62	39.80	39.97	40.15
1.3	40.32	40.49	40.66	40.82	40.99	41.15	41.31	41.47	41.62	41.77
1.4	41.92	42.07	42.22	42.36	42.51	42.65	42.79	42.92	43.06	43.19
1.5	43.32	43.45	43.57	43.70	43.83	43.94	44.06	44.18	44.29	44.41
1.6	44.52	44.63	44.74	44.84	44.95	45.05	45.15	45.25	45.35	45.45
1.7	45.54	45.64	45.73	45.82	45.91	45.99	46.08	46.16	46.25	46.33
1.8	46.41	46.49	46.56	46.64	46.71	46.78	46.86	46.93	46.99	47.06
1.9	47.13	47.19	47.26	47.32	47.38	47.44	47.50	47.56	47.61	47.67
2.0	47.72	47.78	47.83	47.88	47.93	47.98	48.03	48.08	48.12	48.17
2.1	48.21	48.26	48.30	48.34	48.38	48.42	48.46	48.50	48.54	48.57
2.2	48.61	48.64	48.68	48.71	48.75	48.78	48.81	48.84	48.87	48.90
2.3	48.93	48.96	48.98	49.01	49.04	49.06	49.09	49.11	49.13	49.16
2.4	49.18	49.20	49.22	49.25	49.27	49.29	49.31	49.32	49.34	49.36
2.5	49.38	49.40	49.41	49.43	49.45	49.46	49.48	49.49	49.51	49.52
2.6	49.53	49.55	49.56	49.57	49.59	49.60	49.61	49.62	49.63	49.64
2.7	49.65	49.66	49.67	49.68	49.69	49.70	49.71	49.72	49.73	49.74
2.8	49.74	49.75	49.76	49.77	49.77	49.78	49.79	49.79	49.80	49.81
2.9	49.81	49.82	49.82	49.83	49.84	49.84	49.85	49.85	49.86	49.86
3.0	49.87									
3.5	49.98									
4.0	49.997									
5.0	49.99997									

Source: *A Practical Introduction to Measurement and Evaluation* by H. H. Remmers, N. L. Gage, and J. F. Rummel, Harper & Row, 1960, p. 50.

relationships hold true *only* in the case of normal distributions of data as defined by the formula for a normal distribution presented previously.

In a normal distribution of data, the number of cases within the range of one standard deviation above the mean and one standard deviation below the mean is approximately 68 percent of the total. It was shown earlier that the number of cases in the interquartile range was 50 percent of the total. Therefore, in the normal distribution the value of Q is approximately two-thirds the value of the standard deviation, σ. The precise relationship is $Q = .6745\ \sigma$.

In distributions that closely approximate the normal curve, the total range of scores extends about three standard deviation units above and below the mean. Within the limits of one standard deviation on each side of the mean, approximately 68 percent of the cases will be found; within two sigma lengths, approximately 95 percent of the cases; and over 99 percent of the cases will be found within three sigma lengths. If the distribution does not approximate the normal curve, or if it is skewed, these relationships will vary somewhat.

The total area under a normal curve represents the total frequency of values in a distribution. The exact proportion of cases included between ordinates erected at various distances from the mean, when the distances are expressed as proportions of the standard deviation (x/σ), are shown in Table 9, giving the percentage of total area under the normal curve between the mean ordinate and an ordinate at other given distances.

STANDARD AND SCALED SCORES

The score on any variable becomes more meaningful when it can be compared with the scores of well-identified groups of data. For example, a pupil may receive the following scores on four tests:

Arithmetical computation	37
Arithmetical reasoning	21
Reading comprehension	86
Work-study skills	64

These numbers are called *raw scores.* If a pupil is in the fifth grade, it would be desirable to know what these raw scores mean in telling how adequate he is compared with other fifth-grade pupils. It would be possible to use various measures of central tendency and variability, as previously discussed, or the total possible scores on each of the tests. However, the raw scores on different examinations are seldom comparable, because of differences in number of items from test to test, differences in their difficulty, and the varying distributions of the scores. Two techniques for making scores comparable, both of which are in common use, are the transformation of scores into *percentiles,* as previously discussed, and the transformation into *standard scores.*

Standard scores are obtained by adding or subtracting a constant value to or from all raw scores and multiplying the results by another constant. In this manner all the differences among individuals retain their same relative values and the scores for various tests can be made comparable. In developing standard scores, the mean and standard deviation for a distribution are computed. The distribution is then marked off into standard deviations from the mean, as shown in Figure 43. The deviations above the mean are indicated by a plus sign, the ones below the mean by a minus sign, and the mean is indicated by zero. The distance between each of these standard deviation units is called *one standard score.*

z-Scores

The basic standard score is the *z*-score, which is computed by use of the formula:

$$z = \frac{X - M}{\sigma}$$

where z is the standard score,
X is the obtained raw score,
M is the arithmetical mean of the distribution,
σ is the standard deviation of the distribution of scores.

By substituting the raw scores on the four tests into the above formula, they may be transformed to z-scores as follows. The means and standard deviations shown in the second and third columns, below, are given.

	Raw score	M	σ	z-score
Arithmetical computation	37	40	6	$-.5$
Arithmetical reasoning	21	25	5	$-.8$
Reading comprehension	86	70	10	$+1.6$
Work-study skills	64	60	8	$+.5$

In comparison with the group in which this particular fifth grade pupil is a member, the z-scores indicate that he is slightly below the class average in arithmetic, slightly above the average in work-study skills, and considerably above the average in reading comprehension, as measured by the various tests. The standard scores, then, become more meaningful than the raw scores even though the class means and standard deviations differed among the four tests.

It is sometimes awkward, however, to work with scores that have zero, plus, and minus values even though these scores can simplify interpretations and increase comparability. Also, in some instances there is reason to question the uniformity of the units in the raw score scale, as pointed out by Flanagan.

> If the raw-score scale is distorted, with large units at the low score end due to too few easy items and small units at the high end because of many very difficult items, the standard scores will have precisely the same defects. The standard scores described above should be differentiated from other types of transformed scores . . . [and] suggest the need for some other type of unit. The fact that many distributions in raw score units tend to have a bell-shaped form suggested that perhaps units giving a specified shape to the distribution might be stable and useful.[2]

A detailed discussion of problems involved in interpreting test

[2] John A. Flanagan, "Units, Scores, and Norms." *Educational Measurement* (ed. Everett F. Lindquist). Washington, D.C.: American Council on Education, 1951, p. 723.

scores, the technical problems in establishing scores and norms, and various procedures for equating scores, many of which go beyond the scope of this text, are also presented by Flanagan in the same reference.[3]

T-Scores

One of the common practices in transforming scores to avoid the use of fractional and negative values is to modify the units of a score distribution. Transformed scores that have a mean of 50 and a standard deviation of 10 are called *T*-scores. These were first developed to provide a distribution of reading ability for unselected groups of twelve-year-old children.

In computing *T*-scores, each of the obtained *z*-scores is multiplied by 10 and added to 50. Thus, the scores on the above four tests would yield *T*-scores as follows:

	z-score	*T*-score
Arithmetical computation	$-.5$	45
Arithmetical reasoning	$-.8$	42
Reading comprehension	$+1.6$	66
Work-study skills	$+.8$	58

T-scores are usually distributed with values ranging from 20 to 80, correspondingly ranging from -3 to $+3$ standard deviations, with a mean value of 50. Thus, the *T*-score transformation eliminates the negative numbers and decimals obtained with *z*-scores and facilitates many computations. Several standardized tests have made use of a similar transformation by using arbitrarily different means and standard deviations. The Educational Testing Service, Princeton, New Jersey, publishes and administers the Graduate Record Examination for the selection of students to graduate programs in many institutions of higher education. The Graduate Record Examination scores have been transformed to a distribution with a mean of 500 and a standard deviation of 100. The army General Classification Test, employed for the selection of military per-

[3] *Ibid.,* pp. 695–762.

sonnel, uses a standard score distribution with a mean of 100 and a standard deviation of 20. Several other tests have been developed using other similar standard score distributions. Thus any distribution of raw scores can be transformed to a distribution with any desired standard deviation (the chosen multiplier) and any desired mean (the chosen constant added). The polygon of such a transformed distribution will have exactly the same shape as the polygon of the raw scores.

Stanines

During World War II, psychologists in the Air Force developed another type of standard score system called the *stanine* plan. This plan divides the norm population into nine groups, from whence came the term, stanines. With the exception of stanine 9, at the top of the distribution, and stanine 1, at the bottom, the nine groups are spaced in units equal to one-half of a standard deviation. The distribution has a mean of 5 and a standard deviation of 2. Thus, stanine 5 is defined as including the scores within $\pm.25\sigma$ of the mean; stanines 2, 3, 4 and 6, 7, 8 each include scores within additional $.50\sigma$'s from the mean; and stanines 1 and 9 include all scores which fall below -1.75σ and above $+1.75\sigma$, respectively.

The various relationships among the areas under the normal curve, standard deviations, cumulative percentages, percentile equivalents, and a few standard and scaled scores, and stanines are shown in Figure 44. The several scores illustrated are drawn parallel to the base line of the normal curve. The values on any of these scales may be related to the base line of the normal curve by marking off a vertical line from any value. The interpretation of standard score scales derives from the idea of using the normal curve as the mathematical model.

For the researcher who is interested in more detailed information with respect to these scales, and in additional procedures for transforming raw scores into other scales, reference should also be made to the statistical references and articles listed at the end of this chapter.

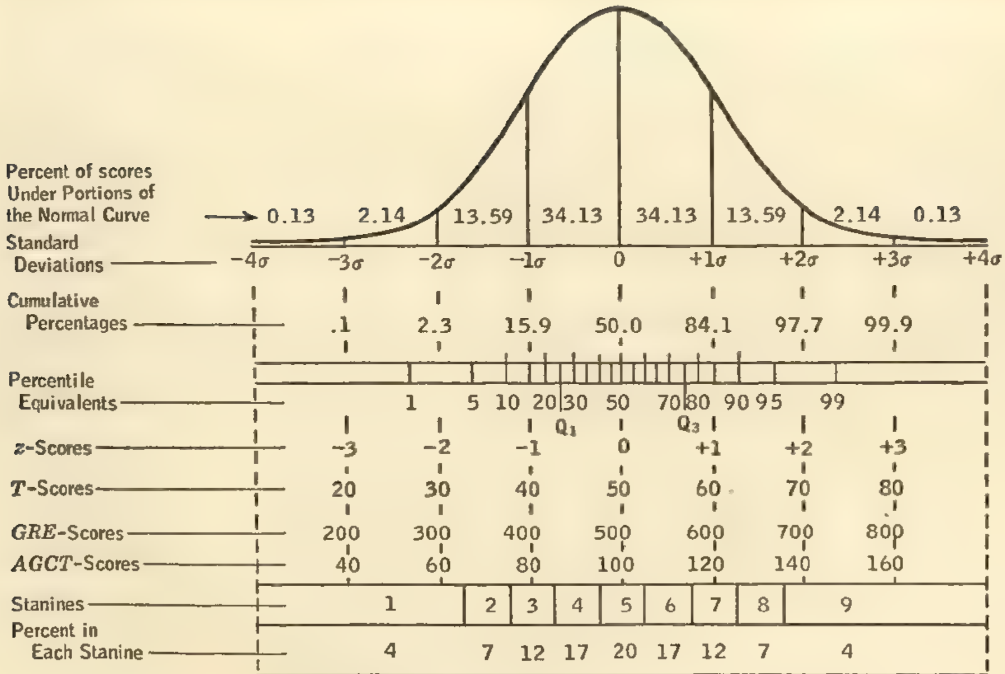


FIGURE 44. The Normal Curve with Corresponding Scaled and Standard Scores.

MEASURES OF RELATIONSHIP

In addition to measures of central tendency and variability for describing characteristics of a single distribution of data, a researcher often finds it necessary and desirable to determine the relationship, or association, between two or more sets of values in two or more distributions. For example, do pupils scoring high on a test in reading receive correspondingly high scores in arithmetic, or spelling, or any other variable? Do individuals far above average in one trait tend to be well above average in another, and those below average in the one tend to be below average in the other? Whenever the measures of each of two traits are secured for each individual and they tend to have about the same relative positions in the distributions of values for the two traits, it may be said that they are "positively related." The *coefficient of correlation* is a single value which is used to represent the relationship between the two sets of data collected for the same individual or for individuals who can be paired in some manner. It represents the extent to which changes in one variable (distribution) are accompanied by changes in another.

Suppose, for example, that the values for a group of 12 individuals are obtained on two variables, X and Y, as shown for Group IV, below. What degree of relationship is there between the two sets of values?

Group IV

Individuals:	A	B	C	D	E	F	G	H	I	J	K	L
Variable X:	10	12	14	16	18	20	20	22	24	26	28	30
Variable Y:	5	6	7	7	9	8	9	10	11	11	13	12

From a general inspection it appears that the Y-values tend to increase as the X-values increase, but that there is not perfect agreement in the two sets of values.

When these data are plotted in a two-way table or *scattergram,* as shown in Figure 45, the relationship between these two sets of values becomes more easily seen. In plotting the

data on a scattergram, a dot is placed at the intersection of the lines corresponding to the values for each individual on each of the two variables. The values on the vertical axis are usually arranged in ascending order from the base line, while those on the horizontal axis are usually arranged in ascending order from the left-hand side of the figure. With this ordering of score

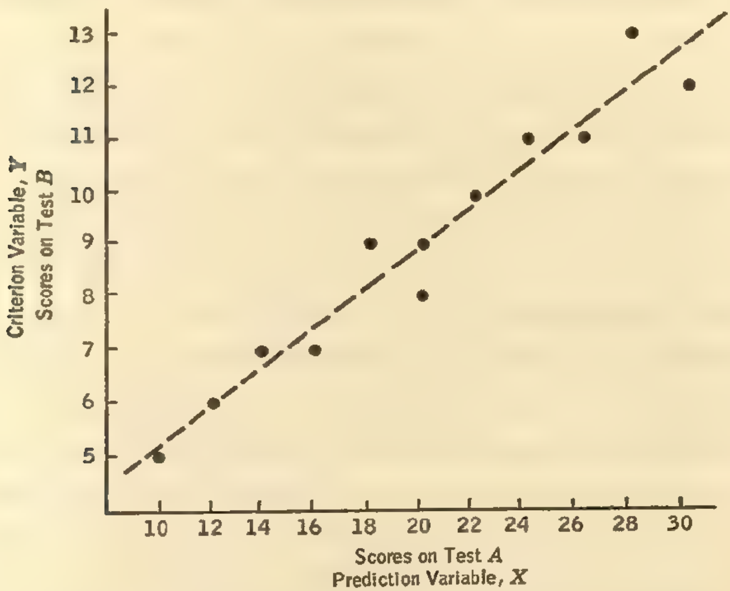


FIGURE 45. Relationship Between Scores on Test A and Scores on Test B for 12 Pupils.

values, the relationship is considered *positive* when the general arrangement of dots on the scattergram forms a pattern running from the lower left to the upper right of the figure. Whenever the dots fall on a straight line, the relationship is linear and perfect, and the *coefficient of correlation* has a value of 1.00.

In some types of data it is found that the values increase on one variable as they decrease on another. In this case the direction of the pattern of the scattergram would run from the upper left to the lower right, and the relationship would be described

as *negative.* The coefficient of correlation is so worked out that when there is complete agreement it has a value of 1.00, which may be indicated as +1.00 if the relationship is positive, or —1.00 if the relationship is negative. *The sign of the coefficient has no meaning other than indicating the direction of the relationship.*

If there is no relationship between the two sets of data, the value of the coefficient of correlation would be zero. This means that there is no systematic increase, or decrease, of values on one set of data corresponding to values on the other set of data. Thus, the range of numerical values of the coefficient of correlation may be defined as extending from a perfect negative relationship (—1.00), to a zero relationship (0.00), to a perfect positive relationship (+1.00). Zero and perfect relationships are rarely, if ever, obtained in actual situations. Usually correlations are found to fall somewhere between the two extremes.

The pattern of values on a scattergram does not always tend to fall along a *straight* line. The line may be curved, in which case the relationship is defined as being *curvilinear.* The four diagrams in Figure 46 illustrate some of the possible relationships in varying degrees. Diagram A shows a relatively *high negative curvilinear* relationship. The values tend to fall along a curved line running in the general direction from the upper left to the lower right to indicate a negative curvilinear relationship. Since these values tend to be close to a hypothetical *line* running through them, the relationship may be described as relatively high. In correlation work this *line* is known as the *regression line.*

Diagram B shows a relatively *low positive relationship.* It is considered low because the values scatter farther out from a hypothetical line, or regression line, than in Diagram A. Diagram C shows a relatively *high positive linear* relationship. Diagram D indicates that the relationship is *practically zero,* or that there is no general trend for the values to fall close to any given line in any direction.

Because of the variety of ways in which variables may be related, there are several indexes of relationship that have been

developed to provide quantitative measures of degrees of relationship. The more commonly used indexes of relationship are the *Pearson Product-Moment Coefficient of Correlation* (r), and the *Rank Order Correlation Coefficient* (*rho* or ρ). The computational procedures for these two indexes are presented in the

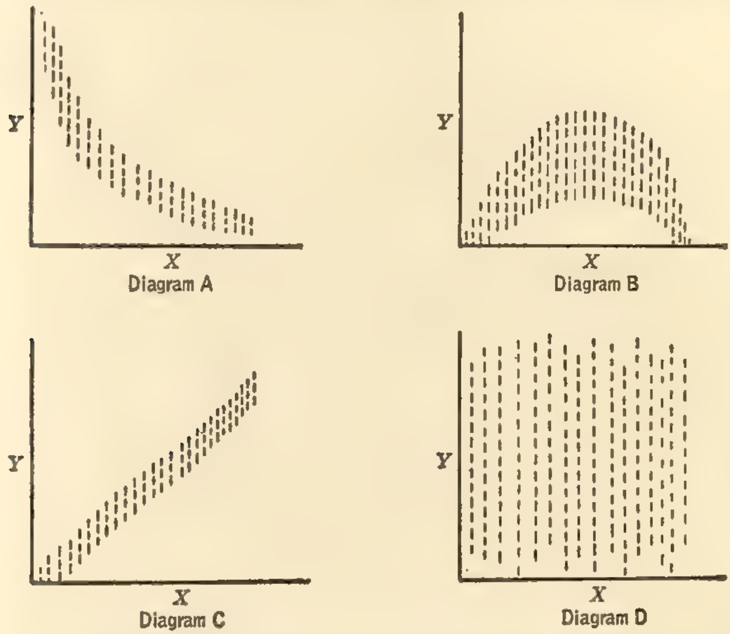


FIGURE 46. Possible Relationships Between Scores on Test X and Test Y.

following sections for the values for individuals in Group IV, above. Other indexes of correlation, for which the reader should refer to standard textbooks on statistical procedures, include the following.

Correlation ratio (eta or η)
Biserial correlation (r_{bis})
Point biserial correlation ($r_{\text{p.bis}}$)
Tetrachoric correlation (r_{tet})
Phi coefficient (ϕ)
Mean square contingency (ϕ^2)

Coefficient of contingency (C)
Chi square (χ^2)
Multiple correlation (R)

Each of these indexes is developed in terms of various characteristics of distributions of data, such as continuous or discrete data, arbitrarily categorized data, and normally distributed data.

Computation of r

The Pearson Product-Moment Coefficient of Correlation (r) may be defined by the formula:

$$r_{xy} = \frac{\Sigma xy}{N\sigma_x\sigma_y}$$

However, for computational convenience this formula may be simplified algebraically to the following which does not require computation of the standard deviations of the variables.

$$r_{xy} = \frac{\Sigma xy}{\sqrt{\Sigma x^2\, \Sigma y^2}}$$

Σxy	is the sum of the products of the paired scores expressed as deviations from their own means
N	is the number of pairs of scores
σ_x, σ_y	are the standard deviations of the respective distributions
Σx^2, Σy^2	are the sum of the squares of the deviation scores in each of the distributions

Table 10 illustrates the form for arranging the data and the computational procedures by use of the *deviation* method. As the number of pairs of values increases, and whenever the means of either distribution do not turn out to be integral values, this procedure becomes very cumbersome and tedious.

While there are several methods of calculating the Pearson Product-Moment Coefficient of Correlation (r),[4] the method

[4] This method is usable only when it is known that the relationship is rectilinear. It may seriously distort the truth when applied otherwise, for example, in the correlation of teachers' salaries and their length of service. For methods of computing correlation coefficients for non-rectilinear relationships, the reader should refer to textbooks on statistical methods.

illustrated here with the use of the *Computing Work Sheet for Correlation,* as shown in Figure 47, and based upon the scores given in Table 11, provides a rapid method by the use of an electric calculating machine. The formula used makes use of the data in raw score form as follows:

$$r_{xy} = \frac{N\Sigma XY - \Sigma X \Sigma Y}{\sqrt{[N\Sigma X^2 - (\Sigma X)^2][N\Sigma Y^2 - (\Sigma Y)^2]}}$$

According to this formula, it is necessary to obtain the following six values from the raw data:

N, ΣX, ΣY, ΣXY, ΣX^2, and ΣY^2.

In using an automatic calculator it is necessary only to enter

TABLE 10. Deviation Method of Computing r

Individuals	Raw Values X	Y	Deviation Values[a] x	y	x^2	y^2	xy
L	30	12	10	3	100	9	30
K	28	13	8	4	64	16	32
J	26	11	6	2	36	4	12
I	24	11	4	2	16	4	8
H	22	10	2	1	4	1	2
G	20	9	0	0	0	0	0
F	20	8	0	−1	0	1	0
E	18	9	−2	0	4	0	0
D	16	7	−4	−2	16	4	8
C	14	7	−6	−2	36	4	12
B	12	6	−8	−3	64	9	24
A	10	5	−10	−4	100	16	40
Sums:	240	108	0	0	440	68	168

$N = 12$ $M_X = 20$ $M_Y = 9$ $\sigma_X = 6.055$ $\sigma_Y = 2.380$

$$r_{xy} = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}} = \frac{168}{\sqrt{(440)(68)}} = \frac{168}{\sqrt{29920}} = \frac{168}{172.974} = .971$$

[a] $x = X - M_x$, $y = Y - M_y$

the pairs of X and Y scores and the machine will yield totals of all the above values with the exception of N. It should be

noted that the ΣXY is given at double value, $2\Sigma XY$, by the calculating machine. Therefore, it is necessary, in using the check sheet for operation number 1, to take half the value shown in the $2\Sigma XY$ column in the "Totals" part of the check sheet. The operations on the check sheet are numbered in the order of their use in determining the correlation coefficient. It should also be noted that the means and standard deviations for each set of scores may be obtained easily in this computational procedure.

X-variable Reading test scores

Y-variable English test scores

Source of Data High School Seniors

Date 2/21/55 Calculator Mr. McBridger Checker

I. Totals from Calculator

N	ΣX	ΣY	ΣX^2	$2\Sigma XY$	ΣY^2
10	418	487	24004	49114	30611
10	653	666	50091	98716	49342
10	537	537	39987	82306	42805
30	1608	1690	114082	230136	122758

II. Operations with Totals

Operation	Calculation	Check
① $N\Sigma XY - \Sigma X \Sigma Y$	734520	
② $\sqrt{N\Sigma X^2 - (\Sigma X)^2}$	$\sqrt{836796} = 914.7655$	
③ $\sqrt{N\Sigma Y^2 - (\Sigma Y)^2}$	$\sqrt{826640} = 909.1974$	
④ $r = ① \div ② \div ③$	.88315	
⑤ $M_x = \Sigma X \div N$	53.60	
⑥ $M_y = \Sigma Y \div N$	56.33	
⑦ $\sigma_x = ② \div N$	30.49	
⑧ $\sigma_y = ③ \div N$	30.31	

FIGURE 47. Computing Work Sheet for Correlation.

TABLE 11. Correlation Between Scores on a Reading Test and an English Test Given to 30 High School Seniors

Pupils	Reading X	English Y	Pupils	Reading X	English Y
1	85	81	16	55	72
2	39	6	17	95	91
3	89	73	18	80	77
4	53	61	19	92	88
5	30	62	20	43	47
6	42	80	21	22	11
7	14	24	22	99	97
8	28	52	23	73	85
9	10	36	24	3	4
10	28	12	25	66	62
11	80	75	26	70	79
12	82	77	27	83	85
13	86	82	28	87	92
14	18	14	29	23	14
15	22	39	30	11	8

$r_{xy} = .883$

$M_x = 53.60$ $\qquad$ $M_y = 56.33$

$\sigma_x = 30.49$ $\qquad$ $\sigma_y = 30.31$

COMPUTATION OF RHO (ρ)

The Rank-Order Coefficient of Correlation may be obtained by solving the following equation.

$$\rho = 1 - \frac{6\Sigma D^2}{N(N^2 - 1)}$$

ΣD^2 is the sum of the squared differences in ranks of the paired values

N is the number of pairs, or individuals

It should be noted that the formula for *rho* includes a term, the differences in ranks (D^2), which is subtracted from 1. If the individuals were ranked in the same order on variable X as on variable Y there would be no differences in ranks, and the sum of the differences squared would be zero. Thus, *rho* would equal 1—a perfect positive correlation. As the differences between the rank orders of these individuals increase there would

be an increased amount subtracted from 1 to reduce the correlation coefficient.

The procedure for the computation of *rho* is illustrated in Table 12. It should also be noted that there are some instances in these distributions where two individuals have the same raw values. Whenever two or more individuals have the same raw value in the original distributions, the ranks which their scores would occupy are averaged and the mean rank value is then assigned to each. For example, individuals I and J, each of whom has a value of 11 on distribution Y, would occupy ranks 3 and 4. Since one should not have a higher rank than the other, the two ranks are averaged arithmetically and the rank of 3.5 is assigned to each. Other tied values are treated in the same manner.

TABLE 12. Method of Computing ρ

Individuals	Raw Scores X	Raw Scores Y	Ranks X	Ranks Y	Difference in Ranks D	Differences Squared D^2
L	30	12	1	2	1.0	1.00
K	28	13	2	1	1.0	1.00
J	26	11	3	3.5	.5	.25
I	24	11	4	3.5	.5	.25
H	22	10	5	5	.0	.00
G	20	9	6.5	6.5	.0	.00
F	20	8	6.5	8	1.5	2.25
E	18	9	8	6.5	1.5	2.25
D	16	7	9	9.5	.5	.25
C	14	7	10	9.5	.5	.25
B	12	6	11	11	.0	.00
A	10	5	12	12	.0	.00
						$\Sigma D^2 = 7.50$

$$\rho = 1 - \frac{6\Sigma D^2}{N(N^2-1)} = 1 - \frac{6(7.5)}{12(144-1)} = 1 - \frac{45}{1716}$$

$$\rho = 1 - .026 = .974$$

A basic difference between the Pearson Product-Moment and the Rank-Order Coefficients is that the Pearson r makes use of the magnitude of values as they actually appear in the distributions, while the Rank-Order *rho* disregards the numerical values

and considers only their corresponding ranks. However, in most practical applications, the value of *rho* usually differs but little from that obtained from the computation of r.

Use of Correlation Coefficients in Prediction

In many research studies problems arise in which it is desirable to predict one characteristic of an individual from one or more other characteristics. For example, since it is known that there is a high positive degree of relationship between two distributions, as shown for the twelve individuals on distributions X and Y in Table 11, one would expect that a value on variable Y could be predicted quite accurately for a *thirteenth* individual by knowing only the value he would make on variable X. That is, a sample of individuals could be used to determine the relationship between two variables, and predictions could then be made for others not in the sample.

In making any predictions, one set of values would be designated as the *criterion* (variable X in the previous example) and the other set of values would be designated as the *predicted* variable (variable Y). If the data representing the criterion and the predicted variables were plotted on a scattergram the line assumed to be satisfactory for predicting would be designated as the *regression line.* The formula for making the predictions would be designated as the *regression equation.*

If there is a perfect relationship between the X and Y variables, the relative position of one individual with respect to others in the distribution would be the same for both sets of data. However, since perfect agreement is rarely obtained, the predicted position of an individual on the Y distribution would tend to be closer to the mean of the Y distribution than his position in the X distribution is to the mean of that distribution. For example, in a distribution of values on a given criterion (X), as shown in Figure 48, there may be several individuals having a given value (indicated by X in the figure). When the Y-variable distribution is made, it may be seen that some of the individuals having a given value (X) would have values (Y) at different points in the distribution of the predicted variable.

These are shown in the figure as arrows fanning out from the X-value to the base line of the Y-values. The best prediction for any one of these individuals having a given X-value would be the mean of the Y-values for the same individuals. If the re-

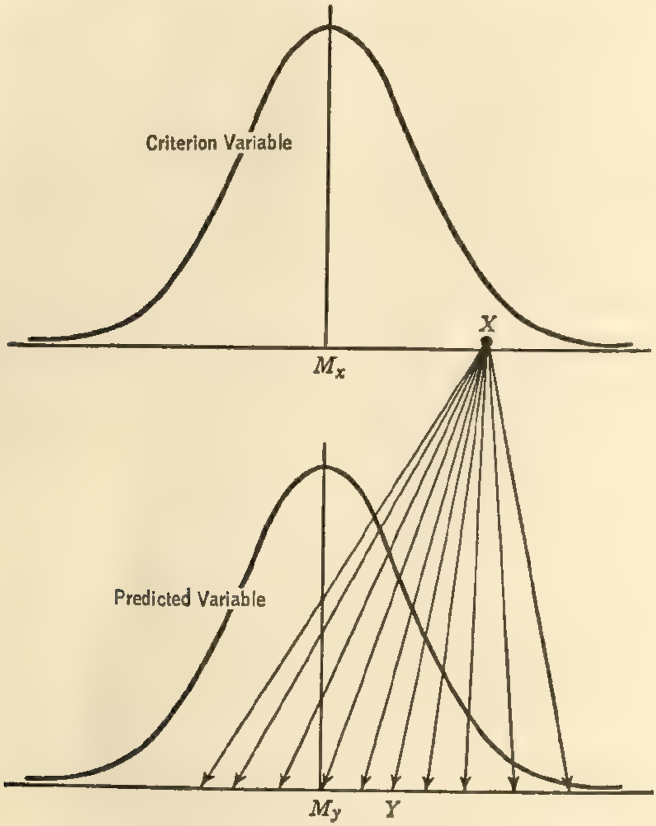


FIGURE 48. The Regression Tendency.

lationship between the X and Y distributions were perfect, all Y-values would fall at the same point on the distribution directly below the X-value on the criterion variable, and the values of the X and Y values could be determined by the following equation:

$$\frac{Y - M_y}{\sigma_y} = \frac{X - M_x}{\sigma_x}$$

In predicting the Y-value for any given X-value, the above equation would be transposed algebraically to yield the value for Y:

$$Y = \frac{\sigma_y}{\sigma_x}(X - M_x) + M_y$$

When the relationship between the two variables is not perfect, the previous equation is modified by the use of the correlation coefficient as follows, and is called the *regression equation:*

$$Y = r_{xy}\frac{\sigma_y}{\sigma_x}(X - M_x) + M_y$$

Consider, for example, how this regression equation may be used if one wished to predict the value that would be made by an individual on variable Y for whom only his value on variable X is available. As shown in Table 11, the following measures were determined on the basis of the twelve individuals who have values on both variables X and Y:

$M_x = 20$	$\sigma_x = 6.055$	$r_{xy} = .97$
$M_y = \ \ 9$	$\sigma_y = 2.380$	

Assume that a *thirteenth* individual had a value of 23 on the X-variable, but that no value was available for him on variable Y. It would be possible to predict the value he would obtain on variable Y by substituting the given values in the previous equation:

$$Y = (.97)\,\frac{2.380}{6.055}\,(23 - 20) + 9 = 10.14$$

When there are several values of X for which predicted values for Y are desired, the formula may be reduced to a general expression and simplified, using the given values, as follows:

$$Y = (.97)\frac{2.380}{6.055}(X - 20) + 9$$

$$Y = .3813X + 1.374$$

By using the latter expression, then, it would be a relatively simple matter to determine the predicted values of Y for all possible values of X. It should be noted that the predicted value for the mean of Y is exactly the same as obtained for the mean of Y. By substituting the mean value of X (20) into the latter formula, the predicted value for Y is 9.

This latter, or general formula, is also the equation of the *regression line*,[5] which is the "best fitting" line for the pattern of dots as shown previously in Figure 45. When the expression, "best fitting" line, is used, it means that the line is chosen so that the sum of the squares of the vertical distances from the line to every point is a minimum. A rough approximation of this line may be made by calculating the mean Y-value for each X-value and drawing a straight line through, or as close as possible to, these means.

Cautions in the Use and Interpretation of Correlation Coefficients

A correlation coefficient represents the extent to which changes in one variable are accompanied by changes in another, or the degree to which the data, when plotted on a scattergram, fall along a line (regression line). One of the most important uses for a correlation coefficient is that of indicating the extent to which values of one variable can be predicted from known values of another variable. One of the important characteristics of a correlation coefficient is that it may be used for making comparisons between variables expressed in different units, such as annual income of individuals and the number of years of formal education they have completed. While correlation coefficients play an important role in research, those who lack training in statistical procedures used in research sometimes misconceive and misuse use them.

One of the common misconceptions is that a correlation coefficient may be interpreted as a percentage. While it has a

[5] It should be noted also, that in any scattergram there are two regression lines, depending upon which variable is called criterion and which predictor. In the above case it would be possible to predict values for X from the several values of Y by transposing the equation algebraically; *i.e.*, $X = 2.6226Y - 3.6035$.

range from zero to unity, as in percentages, *it is in no way a percentage of anything.* It does *not* represent the percentage of relationship existing between two variables, nor does it represent the percentage of accuracy with which one variable may be predicted from another. It is merely an *index*. It is a convenient index of relationship, but it should *not* be considered directly proportional to the degree of relationship. A coefficient of .60 does *not* represent three times as strong a relationship as a coefficient of .20. Nor is the difference in the degree of relationship between r's of .30 and .40 equal to the difference in the degree of relationship between r's of .80 and .90. On the contrary there is likely to be a much larger difference in the degree of relationship between the latter than the former. Figure 49 illustrates the percentage of improvement in prediction over a "best guess" for increasing values of r. This figure shows that

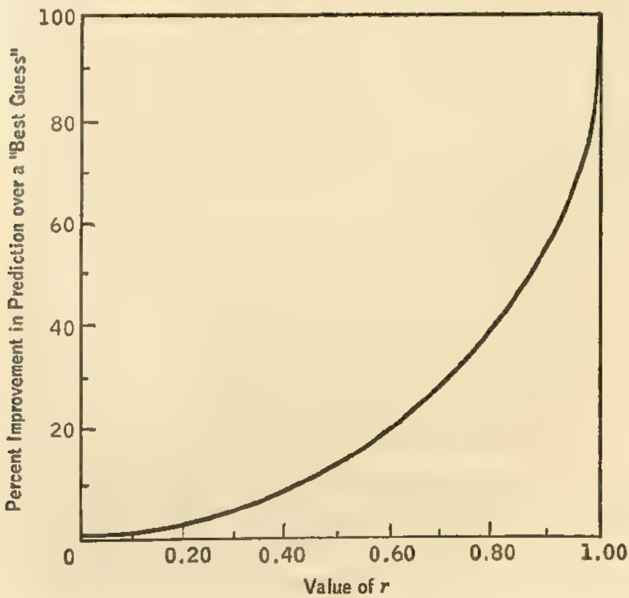


FIGURE 49. Improvement in Accuracy of Prediction for Increasing Values of r. (Reproduced from Everet F. Lindquist, *A First Course in Statistics.* Boston: Houghton Mifflin, 1942, p. 202.)

as the value of r increases the percentage of improvement over a sheer guess increases more and more rapidly.

Another of the common errors in the use of correlation coefficients is the tendency to judge them as "high," "medium," "marked," "low," etc. This tendency is open to serious question since the size of a correlation coefficient can scarcely be considered apart from the purpose for which it is computed. For example, an r of .50 between the number of miles cars are driven and the number of gallons of gasoline consumed would be considered "extremely low," while the same correlation coefficient between measures of bodily growth and the attendance of individuals at civic music concerts would be considered "exceptionally high." The relative importance of the size of correlation coefficients is dependent upon the number of common elements in the two variables being correlated. In the former case there are many common elements with respect to automobiles such that the more miles they are driven the greater is the amount of gasoline needed to propel them, while it would be difficult to find many common elements between the size of individuals and their attendance at musical functions. It should be remembered that correlation coefficients have no *inherent* values in themselves.

Another serious misconception is that of believing that a correlation coefficient indicates a *cause-and-effect relationship.* The causes of relationships are not made explicit in the relationships themselves, but only in the logical reasoning with respect to the characteristics of the variables that are correlated. It may be entirely possible over a period of years to obtain a positive correlation coefficient between the number of failures in college courses and the number of engagements announced by college coeds. However, it is scarcely reasonable to believe that failure in college courses *causes* students to become engaged, or that becoming engaged causes failures in college courses, although it is highly conceivable that some college professors might attribute a causal relationship between the love life of students and college grades. Sometimes "pressure groups" capitalize on the causal-casual misconception of rela-

tionships to "mislead" the public through the use of brochures distributed from door to door. Many times these brochures are "anonymous" and lead one to question their purpose. For example, the writer received a copy of the brochure reproduced in Figures 50 and 51, showing the comparison between the annual per capita consumption of soft drinks and the incidence of polio from 1938 to 1951. The correlation coefficient between the values for these years is approximately .75. The ingenious format, showing a strong pictorial comparison, gives one the impression that the relationship is very high and positive. However, there has been no "proof" that the increase in the consumption of soft drinks *causes* an increase in polio. In fact, during the same period of years, there has been a correspondingly high increase in the number of miles of highway freeways constructed and the number of babies born. Could one say that such an increase in highways was the cause of more polio, or that the drinking of more soft drinks produced a greater number of babies?

TESTING SIGNIFICANCES OF DIFFERENCES

Whenever two measures of central tendency are compared there are always two questions that should arise.

1. Is the difference noted so *small* that it might have resulted from individual variations among the cases drawn for the two samples from which the data were received?
2. Is the difference so *large* that it is unreasonable to expect that individual variations among the cases drawn for the two samples could account for all the difference?

The answers to these questions may be obtained by *tests of significance* expressed in terms of probability. The general procedure for all tests of significance of differences are as follows.

1. Determine the *magnitude of the difference* in measures of central tendency. The difference in the means obtained from the samples is the most frequently used value.
2. Determine the *standard error of the difference* among these values. The standard error of the difference is the measure used

COMPARE!

(1) Annual Per Capita Consumption of Bottled SOFT DRINKS	YEAR	(2) Yearly Incidence of POLIO
[illegible]	1938	[illegible]
[illegible]	1939	[illegible]
[illegible]	1940	[illegible]
133.8	1941	9,086
126.2	1942	4,033
138.6	1943	12,450
147.1	1944	19,029
132.9	1945	13,619
132.3	1946	25,698
150.9	1947	10,734
164.4	1948	27,902
162.0	1949	42,366
158.0	1950	33,351
Est. 173.8	1951	Est. 28,500

Increase or decrease in the incidence of Polio is reflected in the year following a raising or lowering in the per capita consumption of soft drinks.

(1) American Bottlers of Carbonated Beverages, Washington, D. C. (Revised Sept. 1951)

(2) The National Foundation for Infantile Paralysis, New York, N. Y.

FIGURE 50. Comparison of Consumption of Soft Drinks and Incidence of Polio. (Front of anonymous brochure; source unknown.)

PSS-SST

So few people EVEN CHEMISTS realize the TERRIBLE phosphoric acid content of "COLA" drinks.

THIS WILL TRULY OPEN YOUR EYES

Your local producer of

CEMETERY MEMORIALS

will gladly GIVE you a handful of small marble "CHIPS" . . .

100 TIMES HARDER THAN TOOTH ENAMEL!

Quickly insert at least 5 or 6 marble "chips" into a bottle of "cola" . . . any child or adult witnessing this INSTANTANEOUS attack will NEVER drink another damnific "cola" . . .

try this ACID TEST!!

FIGURE 51. Improperly Made Cause-and-Effect Relationships. (Back of anonymous brochure; source unknown.)

to estimate the extent to which differences obtained from samples of the size used would be expected to occur by chance alone.

3. Determine the *ratio of the obtained differences to the standard error of the difference.* This ratio has been designated as the "critical ratio" or "significance ratio" and is usually expressed in "sigma" distances.
4. Determine the *sampling distribution of the differences* and compute the sigma distance of the significance ratio.
5. Determine the *area under the curve* of the sampling distribution of the plus and minus sigma distances. This area is used to determine the probability that the obtained difference in obtained measures of central tendency is no greater than random sampling would suggest.

This procedure, however, is dependent upon the adequacy and precision of the statistical methods used, and needs further consideration with regard to the definition of the hypothesis, methods of stating the hypothesis, choice of the criterion to be used, types of populations, the statistical design, the mathematical models, the confidence level required for rejecting a hypothesis, and the standards required for proof. Since the various techniques for testing a hypothesis are so extensive that they warrant specialized study of statistical methods, only a very elementary introduction to these techniques is presented here. For specific details of methodology, the reader should refer to standard textbooks on statistical procedures.

As a basis for considering tests of significance of difference, assume that it is wished to determine which of two methods of instruction in ninth grade general science is more likely to result in the greater achievement of pupils as indicated by performance on a general science achievement test. While it would be necessary in any good research study to conduct an experiment for the solution of a problem of this nature in several schools, using several teachers and a sizable number of pupils, for purposes of illustration, with only a small amount of mathematical computation, the problem is carried out here with data from only 20 pupils per class in each method in one

school. It is necessary, at this time, to assume that the two classes were comparable with respect to initial academic achievement in general science, to assume that the two methods had been used for a period of time, and to assume that the following data are the scores on a final test of general science achievement. The identification of the individuals making each of the scores below is not important in the solution of the problem, so that only the scores are given.

Test Scores for Two Methods in General Science

Method *A*					Method *B*				
18	24	50	44	61	96	87	81	78	38
40	50	46	74	50	68	58	73	73	43
41	51	74	19	62	51	57	76	74	48
48	30	22	38	28	88	65	47	56	83

The *t*-test

To determine whether one method is better than the other, it is necessary first of all to determine whether or not there is any *real* difference between them. If there is a real difference, one that is greater than could be accounted for by chance factors, it will be possible to determine by an inspection of the measures of central tendency (means) which is the better method. The hypothesis, then, that should be stated is that there is *no difference between the population means of the two methods.* If this hypothesis can be rejected, the only alternative conclusion would be that there is a real difference between the "true" means. If so, one method would be presumed to be better than the other and it would usually be considered to be the one producing the more favorable mean score.

Since these two methods are administered to two different classes of 20 pupils, the two groups would be considered as being "independent random samples" of the population of ninth grade general science pupils of a particular school. Therefore, the test of significance of difference between the means of

the two groups could be made by the t-test using the following formula:

$$t = \frac{M_a - M_b}{\sqrt{\left(\frac{\Sigma x_a^2 + \Sigma x_b^2}{N_a + N_b - 2}\right)\left(\frac{1}{N_a} + \frac{1}{N_b}\right)}}$$

where t is the significance ratio which is used to determine the probability of the obtained difference being larger than chance, by use of tables of t for various degrees of freedom,

M_a, M_b are the means of the two distributions of scores,

N_a, N_b are the numbers of cases or individuals in each group of pupils, and

Σx_a^2, Σx_b^2 are the sums of the squared deviations from the means of the two distributions.

Level of Significance

Prior to making the test of significance it is necessary to determine in quantitative terms what degree of confidence might be placed in the assertion that there is no difference between the population means of the two groups. This is indicated by the *level of significance*. Whenever sampling is used as a device to learn about a large group from an inspection of a smaller group, there is always the possibility that the results from any one sample may be quite different from the values of the total group or population. The elements of "chance" are always present and all inferences about a population must be made in terms of probability. Many researchers have failed to distinguish between those elements which are certain and those which are matters of chance. A distinction between them has been given by Snedecor[6] in the following statement.

> The possession of an attribute by an individual is a fact that may be determined, but the selection of this individual in a sample is a chance event. Often heard is some inaccurate reporting as, "The probability that an Iowa family owns a car is 90%." An Iowa family

[6] George W. Snedecor, *Statistical Methods*, 4th ed. Ames, Iowa: The Iowa State College Press, 1946, p. 7.

either owns a car or doesn't, the accurate phraseology being, "The probability is 90% that a randomly selected Iowa family shall be found to own a car." It is the selection of the family that is the chance event, not the ownership of the car.

If a researcher should conduct the previous experiment a large number of times in the same school with the same size samples, but using different pupils, he would expect to obtain differences in the sample means that varied somewhat from this one particular sample. If he were to conclude that the difference in means for this one sample was large enough to indicate a real difference in the "true" means, the probability is that he would be wrong a certain percentage of the time. The level of significance indicates the probability that an inference or conclusion would likely be wrong in the long run when it is based upon the results of an experiment. When the consequences of being wrong are not serious, it is the usual procedure for tests of significance to be considered at the 5 percent level. This means that a researcher could expect to be wrong in his inferences 5 percent of the time due to chance fluctuations in sampling. The 1 percent level of significance is more often required, however, for research in which the consequences are important.

Degrees of Freedom

The size of t, or the significance ratio, that is necessary to determine whether or not an obtained difference in means is larger than could be expected by chance in terms of the number of cases in the sample is obtained from a table of *t-values* for various degrees of freedom (see Table 13). If the sample is small, the probability that a given difference in means could be attributed to chance factors is greater than if the sample is large. The larger the sample, the less likely differences in means could be attributed to chance factors.

Degrees of freedom are determined by the size of the sample involved and indicate the value of t that should be used in determining the significance of differences in means. The number of degrees of freedom for the t in the previous example is

$(N_a + N_b - 2)$, or 38. For 38 degrees of freedom, the value of t at the 5 percent level of confidence is 2.04. If the t-ratio obtained is in excess of 2.04, the inference is that there is a real difference between the "true" means of the two methods groups.

To compute the t for the previous example, it is necessary to determine the value of the means, the sums of squared deviations of all scores from their respective means, and to know the number of scores for each method. The following values were obtained:

$$M_a = 44 \qquad N_a = 20 \qquad \Sigma x_a^2 = 4988$$
$$M_b = 67 \qquad N_b = 20 \qquad \Sigma x_b^2 = 5198$$

By substituting these values in the formula for t, as shown, the value of t is -4.442.

$$t = \frac{44 - 67}{\sqrt{\left(\frac{4988 + 5198}{20 + 20 - 2}\right)\left(\frac{1}{20} + \frac{1}{20}\right)}} = -4.442$$

Since this value is far in excess of the 2.04 that was determined as necessary to indicate that the obtained difference was in excess of chance, the hypothesis that the means are equal is rejected at a much higher level of significance than 5 percent. Thus, one may conclude that there is a real difference in the means of the two methods and, since the mean for method B was larger than that for method A, one might infer that method B was the better of the two methods. In making this inference, however, it must be remembered that either a real difference exists between the two methods or that the difference obtained from this sample is one that would happen only rarely (less than 5 percent of the time) if there were no real differences between the methods. It should also be remembered that this is merely an illustrative example of the procedures for testing the significance of differences between means and that there are many other procedures that may be used with specific types of data and for various purposes. No researcher should attempt any statistical analysis of data without first determining whether

or not the intended statistical analysis is appropriate for his problem.

TABLE 13. Values of t at the 5 Percent and 1 Percent Levels of Significance

$d.f.$	$t_{.05}$	$t_{.01}$	$d.f.$	$t_{.05}$	$t_{.01}$
1	12.706	63.657	21	2.080	2.831
2	4.303	9.925	22	2.074	2.819
3	3.182	5.841	23	2.069	2.807
4	2.776	4.604	24	2.064	2.797
5	2.571	4.032	25	2.060	2.787
6	2.447	3.707	26	2.056	2.779
7	2.365	3.499	27	2.052	2.771
8	2.306	3.355	28	2.048	2.763
9	2.262	3.250	29	2.045	2.756
10	2.228	3.169	30	2.042	2.750
11	2.201	3.106	40	2.021	2.704
12	2.179	3.055	60	2.000	2.660
13	2.160	3.012	120	1.980	2.617
14	2.145	2.977	∞	1.960	2.576
15	2.131	2.947			
16	2.120	2.921			
17	2.110	2.898			
18	2.101	2.878			
19	2.093	2.861			
20	2.086	2.845			

INDEX